

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 2, 2004, 19:09:19 ; Search time 3550 Seconds
(without alignments)
10659.738 Million cell updates/sec

Title: US-10-029-180-7
Perfect score: 1557
Sequence: 1 atggatcctagaacacatcc.....ccgttcagcccgccagacaa 1557

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_estfun:*
15: em_estom:*
16: em_gss_hum:*
17: em_gss_inv:*
18: em_gss_pln:*
19: em_gss_vrt:*
20: em_gss_fun:*
21: em_gss_mam:*
22: em_gss_mus:*
23: em_gss_pro:*
24: em_gss_rod:*
25: em_gss_phg:*
26: em_gss_vrl:*
27: gb_gss1:*
28: gb_gss2:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	380.4	24.4	442	9 AA785001	AA785001 g4b03al.r
2	171.6	11.0	768	12 BM077105	BM077105 TrEST-A34
3	108.8	7.0	687	10 BF251517	BF251517 EST418778
4	82.8	5.3	782	28 AQ161501	AQ161501 mgxb00008A

5	5.3	585	28	AQ255041	AQ255041 mgxb00008J
6	3.6	807	13	BX464554	BX464554 BX464554
7	3.5	1448	29	CC220110	CC220110 CH261-92F
8	3.4	1201	9	AL515475	AL515475 AL515475
9	3.4	727	13	BX319543	BX319543 BX319543
10	3.3	1137	29	CC315110	CC315110 TAM32-30A
11	3.2	567	28	AZ895651	AZ895651 RPCI-24-2
12	3.2	1201	13	BX457568	BX457568 BX457568
13	3.2	1201	13	BX421672	BX421672 BX421672
14	3.2	802	29	CNS01170	AL100686 Drosophil
15	3.2	786	28	AQ745021	AQ745021 HS_5507_A
16	3.2	897	29	AG060530	AG060530 Pan trogl
17	3.2	915	13	BQ952432	BQ952432 AGENCOURT
18	3.2	1262	12	BM560943	BM560943 AGENCOURT
19	3.2	1287	10	BF256624	BF256624 HVSMEF001
20	3.1	481	9	AA695157	AA695157 GM02329.5
21	3.1	759	13	BU314584	BU314584 603543708
22	3.1	834	28	BZ051898	BZ051898 jnr57e08
23	3.1	1101	29	CNS00008	AL051971 Drosophil
24	3.1	1201	9	AL514195	AL514195 AL514195
25	3.1	458	10	BF063108	BF063108 7h83e05.x
26	3.1	648	12	BI227246	BI227246 602949793
27	3.1	1026	29	AG081192	AG081192 Pan trogl
28	3.1	732	29	BX149729	BX149729 Danio rer
29	3.1	815	29	BX239144	BX239144 Danio rer
30	3.1	987	29	CNS00418	AL066537 Drosophil
31	3.1	1137	29	AG078502	AG078502 Pan trogl
32	3.1	786	28	BH018846	BH018846 L2011a.d
33	3.1	2152	29	CC283954	CC283954 CH261-181
34	3.1	785	29	AG081505	AG081505 Pan trogl
35	3.1	1130	29	AG060387	AG060387 Pan trogl
36	3.1	1201	13	BX356664	BX356664 BX356664
37	3.1	759	29	AG126157	AG126157 Pan trogl
38	3.1	774	29	AG128958	AG128958 Pan trogl
39	3.1	887	29	CNS025C6	AL211695 Tetraodon
40	3.1	1082	9	AL574528	AL574528 AL574528
41	3.1	1104	13	BQ671688	BQ671688 AGENCOURT
42	3.1	1122	29	CC231036	CC231036 CH261-27F
43	3.1	731	29	CNS042Q8	AL271817 Tetraodon
44	3.1	789	14	CA320647	CA320647 UI-M-FW0-
45	3.1	1201	13	BX365323	BX365323 BX365323

ALIGNMENTS

RESULT 1
AA785001
LOCUS
DEFINITION
AA785001 442 bp mRNA linear EST 29-JUL-1998
g4b03al.r1 Aspergillus nidulans 24hr asexual developmental and
vegetative CDNA lambda zap library Emericella nidulans cDNA clone
g4b03al 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AA785001
g4b03al.r1
vegetative CDNA lambda zap library Emericella nidulans cDNA clone
g4b03al 5', mRNA sequence.
AA785001
AA785001.1
EST.
Emericella nidulans (anamorph: Aspergillus nidulans)
Emericella nidulans
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; Emericella.
1 (bases 1 to 442)
Kupfer,D., Gray,J., Hausner,J., Lai,H., Martin,W., Aramayo,R.,
Prade,R. and Roe,B.
An Aspergillus nidulans EST Database
Unpublished
Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center

QY	512	GAAAGCCTCAAGGTTGGGACCCCGTTGCTG-----CGAATGGTGTGCGCGTATCCCG	565
Db	411	ACAAGCAACCCGTATGGGCTGGCGCAGAGGACTGCCACCCACCTCTGGCCCTACATGC	470
QY	566	GGAAGGACTCCCCCGAACCAGGTTGTGTTCTCAGGGGCGACGCGGTATCCTTCCGA	625
Db	471	CAAAAGATCTCCGAGGACACAAGTGGTAGGAACCCAGGCGCGAGAGGTATCTTCCGA	530
QY	626	GTGTTCCGGGACGCGCAACTCCGGTCAAAATGGTGTAAACGGCACCGGCAAGAACACTA	685
Db	531	GTGTACCCGGTCCGGCTGC---GGTTACAACGGTGTGAATGGCAC---CAAGGGAACAA	583
QY	686	CTATCCCGGCCAAGGATGCCGA-TGGAAGTTCCCTTGCCCGCACTGTAAACAAGACTTAT	744
Db	584	CCATCCCTGCCAAGGATGCAGACGCGGAAAGTTTCCCTGTCCGCACTGCAATAAGACTTTA	643
QY	745	CTTCATGCCAAGCATCTCAAGCGCCATCTGCTACGCCACACTGG	788
Db	644	CTGGCATGCGAAGCACCTTAAGCGCATCTCTTACGCCACACCGG	687
RESULT 4			
AQ161501			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
Unpublished			
Contact: Dean RA			
Clemson University Genomics Institute			
Clemson University			
100 Jordan Hall, Clemson University, Clemson, SC 29634			
Tel: 864 656 5737			
Fax: 864 656 4293			
Email: rdean@clemson.edu			
Seq primer: GGAACAGCTATGACCATG			
Class: BAC ends			
High quality sequence stop: 315.			
Location/Qualifiers			
1..782			
/organism="Magnaporthe grisea"			
/mol_type="genomic DNA"			
/strain="70-15"			
/db_xref="taxon:148305"			
/clone="mgxb0008A02r"			
/tissue_type="Protoplasts"			
/lab_host="E. coli DH10B"			
/clone_lib="CUGI Rice Blast BAC Library"			
/note="Vector: pBACWICH; Site_1: HindIII; Site_2: HindIII;			
Rice blast is one of the most devastating fungal diseases			
of rice world wide. It is a filamentous ascomycete with			
a haploid genome (n=7) of approximately 40 Mbp. Rice			
blast is an important model fungal pathogen for studying			
numerous aspects of the fungal-host interaction. In			
order to facilitate genome wide analysis, a BAC library			
containing 9216 clones with an average insert size of 130			
kbp was constructed. This library represents greater			
than 25X genome coverage. High density colony filters			
are available upon request."			
BASE COUNT	172 a	284 c	192 g
ORIGIN	134 t		

Db	674	ATCCGGCGGGCAACCCCAAGGCGCTTGCCCTTTTGGTCCC	717
RESULT 3			
BF251517			
LOCUS	BF251517		
DEFINITION	EST418778 Coccidioides immitis spherule cDNA library Coccidioides immitis cDNA clone CIAAH91 5' sequence, mRNA sequence.		
ACCESSION	BF251517		
VERSION	BF251517.1		
KEYWORDS	EST.		
SOURCE	Coccidioides immitis		
ORGANISM	Coccidioides immitis		
REFERENCE	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Onygenales; mitosporic Onygenales; Coccidioides.		
AUTHORS	1 (bases 1 to 687)		
TITLE	Gardner,M.J. and Kirkland,T.		
JOURNAL	Generation of ESTs from Coccidioides immitis spherule cDNA library		
COMMENT	Unpublished		
Contact: Malcolm J. Gardner			
Department of Eukaryotic Genomics			
The Institute for Genomic Research			
9712 Medical Center Drive, Rockville, MD 20850, USA			
Tel: 301 838 3519			
Fax: 301 838 0208			
Email: gardner@tigr.org.			
Location/Qualifiers			
1..687			
/organism="Coccidioides immitis"			
/mol_type="mRNA"			
/db_xref="taxon:5501"			
/clone="CIAAH91"			
/dev_stage="spherule"			
/lab_host="SOLR"			
/clone_lib="Coccidioides immitis spherule cDNA library"			
/note="Vector: pBluescript SK(-); Site_1: EcoR1; Site_2: Xho1"			
BASE COUNT	167 a	234 c	160 g
ORIGIN	126 t		
Query Match 7.0%; Score 108.8; DB 10; Length 687;			
Best Local Similarity 56.2%; Pred. No. 9.3e-17;			
Matches 396; Conservative 0; Mismatches 252; Indels 56; Gaps 8;			
QY	95	TGCCTCAGTACAGCATGCGCTCAGTACCCAGTCTCTCAGCGGCACACCCCTGCCTCCTC	154
Db	30	TCCACCAATATCAGATGCATTCGCCGTACCGGATGGCGCAACCGCATACGTTGCCGCCAT	89
QY	155	TGCAACCCCATCATAGCCAGTCGCCCGCTCCTCACTCGTACATGGGCGAGCGCGGTACC	214
Db	90	TGCAACACCATCAGAAATCGTCGCCAAATTCAGCATACGTATCTC-----CAGCCTTTCC	143
QY	215	GGCCTGATCTGAACAGGTACCCCGCA---TCAAGTCACGATGTTTACGGCTCTTCTGCTG	271
Db	144	GGAATGACATGCCTAGATACCCGACAACTCCGCGCAGCGATGTTATGCGGTATCGTCGG	203
QY	272	CGCCGATAATGCCCCACACTACCGTGGGCGAGCTTGCCCTCCGACATCTTTCTTCTCATC	331
Db	204	CGCCGTTGACTACGATGCGCCCGTCAACAGTTTGCCACCGTCAACTTTTCTT-----	256
QY	332	CCAATCCGCGAGCGCAGGCAGCGGCAGCAATCGCCGCACTATCCTCCTCTCATAGCG	391
Db	257	-----GGCCACCAACATCCACAGCAATTCAGGCCACATCATA	293
QY	392	TGCTCCCGCCCGCTTCAGCGCTCAGTCGTACCCGAGCCCAATTGGCGCGCGCTCCCC	451
Db	294	TGCTCCCAACCCAGCAGCAGCCACCCAGCATACCCGCAACCAATCGCGCCAGCGCACCGC	353
QY	452	GGGACCGTCGTGCTGACTTCAACAATGGACTTCCTTCAGGAGCATTCAGTTATTCGGACG	511
Db	354	GTGACAGAAGATCAGAGTATGGCGCTG---TGCCGTGGCCCCCGTTTACAACGGGGAGA	410


```
Query Match      5.3%; Score 82.8; DB 28; Length 782;
Best Local Similarity 51.5%; Pred. No. 3.8e-10;
Matches 271; Conservative 0; Mismatches 242; Indels 13; Gaps 3;

QY 321 CCTTCTCATCCCAATCCGAGGCGCAGGCACAGCGCAGCAATCGCGCACTATCCTCC 380
Db 116 CCCGTACCTTGGCATGATGCCGCGCGCGAGTACCAACAACCTCACCCCGGTATCCAC 175

QY 381 TCCTCATAGCGTGCTCCCGCCCGCTTCCAGCGCTCAGTCGTACCCGCGAGCAATTCGCC 440
Db 176 CACAACCTCAATGATACCCAGGCATCGACCGCAATGGCCACCCCTCAACCTATTCGCC 235

QY 441 GCGGCTCCCGGACCGTCTGCTGACTTCAACAATGGACTTCCTTCAGGAGCATTTCAG 500
Db 236 CGCGCTCTCTGGTGGTCCCGCTCCCGGTATTGCGTCAATGCCGCGGTGGCATGGT 295

QY 501 TTATTCGGACGGAAG-----CCTCAAGGTTGGACCCCGTTGCTGCGAATGGTGTGCG 555
Db 296 CCCTCAGGCGGCATGTCTGTCGCATACGCGCAGAGCCCATGATGCCCCAG--CCTCAA 353

QY 556 CCGTATCCCGGAAGGACTCCCCCGAACCAGGTTGTTGGTTCTCAGGGGCGAGCGGT 615
Db 354 ATGATGCAACAGGAAGGCGACGCGCGACACATGTGTGGTCTCTCAGGGTCGCGGAGGA 413

QY 616 ATCCTTCCGAGTGTTCGGGACGCGCAACTCCGGTCAAAATGGTGTAAACGGCAGCGGC 675
Db 414 ATCTTCCGAGCGCGCGGTAGG-----CCCCAGGCGCGGTGCTGCTGGCAGCGCGG 467

QY 676 AAGAACTACTATCCCGGCAAGGATGCGCGATGGAAGTTCCCTTGCCCGCACTGTAAC 735
Db 468 GCAAAGAACCATATCCCCCAGAAGGATGCCGACGGGAAGTTCCTGTCCTCATTTGTAAC 527

QY 736 AAGACTTATCTTCATGCAAGCATCTCAAGCGCCATCTGCTACGCCACACTGTCGACCGC 795
Db 528 AAGACGTACCTGCACGCCAAGCATCTCAAGAGGCATCTCTTGAGACGTAAGTTTGCCT 587

QY 796 CCGTACATGTGTGTTCTTTGCAAGACACCTTCTCTCGCAGTGATA 841
Db 588 GCTTGTGTTCTGACGTGAAGTGTGCTCAACTGACACATTAATA 633

RESULT 5
AQ255041
LOCUS
DEFINITION
mgxb008J11r CUGI Rice Blast BAC Library Magnaporthe grisea genomic
clone mgxb008J11r, genomic survey sequence.
ACCESSION
AQ255041
VERSION
AQ255041.1 GI:3779356
KEYWORDS
GSS.
SOURCE
Magnaporthe grisea (anamorph: Pyricularia grisea)
ORGANISM
Magnaporthe grisea
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthae.
REFERENCE
1 (bases 1 to 585)
AUTHORS
Yu,Y., Zhu,H., Boyd,C.A., Gaudette,B., Gayle,A., Kingsbury,R.,
Phillips,K., Sasinowski,M, Wing,R.A. and Dean,R.A.
TITLE
A BAC End Sequencing Framework to Sequence the Magnaporthe grisea
Genome
JOURNAL
Unpublished
COMMENT
Contact: Dean RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson University, Clemson, SC 29634
Tel: 864 656 5737
Fax: 864 656 4293
Email: rdean@clemson.edu
Seq primer: GGAACAGCTATGACCATG
Class: BAC ends
High quality sequence stop: 236.
location/Qualifiers
1..585
/organism="Magnaporthe grisea"
/mol_type="genomic DNA"
```

```
/strain="70-15"
/db_xref="taxon:148305"
/clone="mgxb008J11r"
/tissue_type="Protoplasts"
/lab_host="E. coli DH10B"
/clone_lib="CUGI Rice Blast BAC Library"
/note="Vector: pBACWICH; Site_1: HindIII; Site_2: HindIII;
Rice blast is one of the most devastating fungal diseases
of rice world wide. It is a filamentous ascomycete with
a haploid genome (n=7) of approximately 40 Mbp. Rice
blast is an important model fungal pathogen for studying
numerous aspects of the fungal-host interaction. In
order to facilitate genome wide analysis, a BAC library
containing 9216 clones with an average insert size of 130
kbp was constructed. This library represents greater
than 25X genome coverage. High density colony filters
are available upon request."

BASE COUNT      127 a      220 c      145 g      93 t
ORIGIN

Query Match      5.3%; Score 82; DB 28; Length 585;
Best Local Similarity 53.2%; Pred. No. 5.3e-10;
Matches 248; Conservative 0; Mismatches 205; Indels 13; Gaps 3;

QY 321 CCTTCTCATCCCAATCCGAGGCGCAGGCACAGCGCAGCAATCGCGCACTATCCTCC 380
Db 116 CCGGTACCTTGGCATGATGCCGCGCGCGAGTACCAACAACCTCACCCCGGTATCCAC 175

QY 381 TCCTCATAGCGTGCTCCCGCCCGCTTCCAGCGCTCAGTCGTACCCGCGAGCAATTCGCC 440
Db 176 CACAACCTCAATGATACCCAGGCATCGACCGCAATGGCCACCCCTCAACCTATTCGCC 235

QY 441 GCGGCTCCCGGACCGTCTGCTGACTTCAACAATGGACTTCCTTCAGGAGCATTTCAG 500
Db 236 CGCGCTCTCTGGTGGTCCCGCTCCCGGTATTGCGTCAATGCCGCGGTGGCATGGT 295

QY 501 TTATTCGGACGGAAG-----CCTCAAGGTTGGACCCCGTTGCTGCGAATGGTGTGCG 555
Db 296 CCCTCAGGCGGCATGTCTGTCGCATACGCGCAGAGCCCATGATGCCCCAG--CCTCAA 353

QY 556 CCGTATCCCGGAAGGACTCCCCCGAACCAGGTTGTTGGTTCTCAGGGGCGAGCGGT 615
Db 354 ATGATGCAACAGGAAGGCGACGCGCGACACATGTGTGGTCTCTCAGGGTCGCGGAGGA 413

QY 616 ATCCTTCCGAGTGTTCGGGACGCGCAACTCCGGTCAAAATGGTGTAAACGGCAGCGGC 675
Db 414 ATCTTCCGAGCGCGCGGTAGG-----CCCCAGGCGCGGTGCTGCTGGCAGCGCGG 467

QY 676 AAGAACTACTATCCCGGCAAGGATGCGCGATGGAAGTTCCCTTGCCCGCACTGTAAC 735
Db 468 GCAAAGAACCATATCCCCCAGAAGGATGCCGACGGGAAGTTCCTGTCCTCATTTGTAAC 527

QY 736 AAGACTTATCTTCATGCAAGCATCTCAAGCGCCATCTGCTACGCCACACTGTCGACCGC 781
Db 528 AAGACGTACCTGCACGCCAAGCATCTCAAGAGGCATCTTCTGAGAC 573

RESULT 6
BX464554
LOCUS
DEFINITION
BX464554 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE007YB11
3-PRIME, mRNA sequence.
ACCESSION
BX464554
VERSION
BX464554.1 GI:31021646
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 807)
AUTHORS
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished
```



```
COMMENT      Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 3370.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS1DE002ZF03NP1&cluster=3370.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS1DE002ZF03NP1.

FEATURES
source
1..807
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DE007YB11"
/tissue_type="PLACENTA"
/clone_lib="Homo sapiens PLACENTA"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."

BASE COUNT      5 a  609 c  15 g  16 t  162 others
ORIGIN
Query Match      3.6%; Score 56.2; DB 13; Length 807;
Best Local Similarity 31.5%; Pred. No. 0.0021;
Matches 178; Conservative 97; Mismatches 290; Indels 0; Gaps 0;

QY 15 CCATCCCTCTCGGCTCCGCTTACAGTCTGCCTCAAGGATCGGCGCTCTTCTCTGC 74
Db 230 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 289

QY 75 TCCCATCTCGAGATGCCAATGCTCAGTACAGATGCAGCTCAGTACCCAGTCTCTCA 134
Db 290 CCCCCVCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 349

QY 135 GCGGCACACCTGCTCTGTGCAACCCCATCATAGCCAGTGGCCGCTCTCTCACTCGTA 194
Db 350 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 409

QY 195 CATGGGGCAGCGCGGTACCGGCTGTATCTGAACAGGTACCCCGCATCAAGTCACGATGT 254
Db 410 SCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 469

QY 255 TTACGCGTCTTCTGTGCGCGGATAATGCCACACACTACCGTGGGAGCTTGCCTCCGAC 314
Db 470 CCCCCSCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 529

QY 315 ATCTTTCTTCTCATCCCAATCCGCGAGGCGCAGGCACAGGCGCAGCAATCGCCGCACTA 374
Db 530 CCCCCSCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 589

QY 375 TCCTCTCTCATAGCGTGTCTCCCGCCCGCTTCCAGGCTCAGTGTACCCGAGCCAAAT 434
Db 590 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 649

QY 435 TGGCGCGGCGCTCCCGGGACCGTGTGTGACTTCAACAATGAAGTCTCTTTCAGGAGC 494
Db 650 YCCCCYCCYKCCYKGBBKYYTCYBKSTGSKGKYGKSGKSGYBCKSSSSSC 709

QY 495 ATTCAAGTATTCGGACGGAAGCCCTCAAGGTTGGGACCCCGTTGCTGCGAATGGTGTGC 554
Db 710 CTYSBKCKTKTKGYSYBTBSCCCYCCYBCYCYCTCSBKSKSGSKBKSICYBYCB 769

QY 555 GCGGTATCCGGGAAGGACTCCCCC 579
Db 770 YBCYYKSTGTCCGCSYKCCSY 794

RESULT 7
CC220110/c
LOCUS      1448 bp DNA linear GSS 12-MAY-2003
DEFINITION CH261-92F21_Sp6.1 CH261 Gallus gallus genomic clone CH261-92F21,
genomic survey sequence.
ACCESSION  CC220110
VERSION     CC220110.1 GI:30543022
KEYWORDS   GSS.
SOURCE     Gallus gallus (chicken)
ORGANISM   Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 1448)
Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J.,
Warren,W., Graves,T., Mardis,E. and Wilson,R.
Gallus gallus BAC End Reads
Unpublished
Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 182000 Std Error: 0.00
Seq primer: Sp6 ATTTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 1182
High quality sequence stop: 1246.

FEATURES
source
1..1448
/organism="Gallus gallus"
/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="CH261-92F21"
/sex="female"
/cell_line="UCD001, inbred 256"
/clone_lib="CH261"
/note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
CH261 Female Chicken library - For library and clone
ordering information: http://www.chori.org/bacpac"

BASE COUNT      32 a  82 c  1229 g  33 t  72 others
ORIGIN
Query Match      3.5%; Score 53.8; DB 29; Length 1448;
Best Local Similarity 45.6%; Pred. No. 0.011;
Matches 190; Conservative 0; Mismatches 227; Indels 0; Gaps 0;

QY 15 CCATCCCTCTCGGCTCCGCTTACAGTCTGCCTCAAGGATCGGCGCTCTTCTCTGC 74
Db 806 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 747

QY 75 TCCCATCTCGAGCATGCCAATGCTCAGTACAGATGCAGCTCAGTACCCAGTCTCTCA 134
Db 746 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 687

QY 135 GCGGCACACCTGCTCTGTGCAACCCCATCATAGCCAGTGGCCGCTCTCTCACTCGTA 194
Db 686 CCCCCCCCCCGCGCGCGCACACCCACCCCTTTGGCCCCCCCCCCCCCCCCCA 627

QY 195 CATGGGGCAGCGCGGTACCGGCTGTATCTGAACAGGTACCCCGCATCAAGTCACGATGT 254
Db 626 CACGGGCGCGCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 567

QY 255 TTACGCGTCTTCTGTGCGCGGATAATGCCCCACACTACCGTGGGAGCTTGCCTCCGAC 314
Db 566 CCCCCCGCGCCCCCCCCCGCCCCCCCCCCCCCCCCCGTCCCGCCCTGCCCCCCC 507

QY 315 ATCTTTCTTCTCATCCCAATCCGAGGCGCAGGCACAGGCGCAGCAATCGCCGCACTA 374
Db 506 CCCCCCCCCCGGTCCGCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCC 447

QY 375 TCCTCTCTCTCATAGCGTGTCTCCCGCCCGCTTCCAGGCTCAGTGTACCCGAGCC 431
```


Db 591 AACGTGAGTCTGAAAGCCACGCTCCAGCGCTACCACCAGGCGCTGGCTACGGGACC 647

CC315110 1137 bp DNA linear GSS 14-MAY-2003
TAM32-30A5_Sp6.1 TAM32 Gallus gallus genomic clone TAM32-30A5,
genomic survey sequence.
CC315110
VERSION CC315110.1 GI:30709165
KEYWORDS GSS.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus

REFERENCE 1 (bases 1 to 1137)
AUTHORS Kremitzki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J.,
Warren, W., Graves, T., Mardis, E. and Wilson, R.
TITLE Gallus gallus BAC End Reads
JOURNAL Unpublished
COMMENT Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 152000 Std Error: 0.00
Seq primer: Sp6 ATTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 155
High quality sequence stop: 352.

FEATURES
source
1..1137
/organism="Gallus gallus"
/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="TAM32-30A5"
/sex="female"
/cell_line="UCD001, inbred 256"
/clone_lib="TAM32"
/note="Vector: pECBAC1; Site 1: EcoRI; Site 2: EcoRI;
TAM32 Female Chicken library - for library and clone
ordering information: http://www.hbz.tamu.edu"

BASE COUNT 189 a 587 c 157 g 178 t 26 others

ORIGIN

Query Match 3.3%; Score 51.4; DB 29; Length 1137;
Best Local Similarity 43.6%; Pred. No. 0.039;
Matches 199; Conservative 0; Mismatches 257; Indels 0; Gaps 0;

QY 20 CCTCTCGGCGCTCCGTTACAGTCTGCTCAAGGATCGGCGCCTCTTCTCTGCTCCCA 79
Db 514 CCACGCACCGCTCGCGGCGCCCGAGCCAGCCCGCGCCCGACACCTCCCGCTCCCGCG 573

QY 80 TCTCGAGCATGCCAATGCCTCAGTACAGATGAGCAGCTCAGTACCCAGTCTCTCAGCCGC 139
Db 574 CCCCACCGCGCGCCACCGCCACCGCCACACACACCGCGCCACCGCCAGCCCGAGCCAG 633

QY 140 ACACCGCTGCTCTCTGCAACCCCATCATAGCAGTCGCGCGCTCTCTCACTCGTACATGG 199
Db 634 GCG 693

QY 200 GGCAGCGCGCGTACCGGCGCTGATCTGAACAGGTACCGCGCATCAAGTACAGATGTTTACG 259
Db 694 CGACCAACCG 753

QY 260 CGTCTTCTGTCGCGCGCATATATGCCCGACACTACCGTGGGAGCTTGGCTCCGACATCTT 319
Db 754 CTTGTTCTCTCTTCCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 813

QY 320 TCCTTTCTCATCCCAATCCGAGGCGCGAGGCGCGAGCAATCGCGCGCACTATCCTC 379
Db 814 GCG 873

QY 380 CTCCTCATAGCGTCTCCCGCGCGCTTCCAGCGGCTCAGTGTACCCGCAATTCGCGC 439
Db 874 ACACCCACCGCGCGCGGACCCACCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 933

QY 440 CGGCGCGCTCCCGGAGCGCTGCTGCTGACTTCAACA 475
Db 934 GGCACCCCGACCGCA 969

RESULT 11
AZ895651/c
LOCUS
DEFINITION
AZ895651 567 bp DNA linear GSS 05-MAR-2001
RPCI-24-211J20.TJ RPCI-24 Mus musculus genomic clone RPCI-24-211J20
, genomic survey sequence.
ACCESSION
VERSION AZ895651.1 GI:13214596
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 567)
Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,
Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E.,
Russell, D., de Jong, P. and Fraser, C.M.
TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished
COMMENT Other GSSs: RPCI-24-211J20.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html). BAC end
plate: 211 row: J column: 20
Seq primer: SP6
Class: BAC ends.

FEATURES
source
1..567
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-211J20"
/sex="Male"
/cell_type="Spleen/Brain"
/clone_lib="RPCI-24"
/note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."

BASE COUNT 189 a 93 c 254 g 31 t

ORIGIN

Query Match 3.2%; Score 50.6; DB 28; Length 567;
Best Local Similarity 53.2%; Pred. No. 0.047;
Matches 107; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 19 CCCTCTCGGCGCTCCGTTACAGTCTGCTCAAGGATCGGCGCGCTCTTCTCTGCTCCC 78
Db 343 CCTCTCTGCTCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 284

QY 79 ATCTCGAGCATGCCAATGCCTCAGTACACGATGCGAGCTCAGTACCCAGTCTCTCAGCGC 138
Db 283 CTGCGCTCTCTGCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 224


```
QY 139 CACACCCCTGCTCTCTGCAACCCCATATAGCCAGTGCCTCCCTCCTCCTACCTGCTACATG 198
Db 223 CTTCTCTGCTCTCTGCTCTCTGCTCTCTGCTCTCTGCTCTCTGCTCTCTGCTCTGCTATG 164

QY 199 GGGCAGCGCGGTACCGGCCT 219
Db 163 CCTCTGCTCTGCTCTGCTCT 143

RESULT 12
BX457568
LOCUS
DEFINITION BX457568 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
CS0DA001YI06 3-PRIME, mRNA sequence.
ACCESSION BX457568
VERSION BX457568.1 GI:31036780
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
was not normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
7353.r For more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DA001BE03NP1&cluster=7353.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DA001BE03NP1.
FEATURES
source
Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DA001YI06"
/tissue_type="NEUROBLASTOMA"
/clone_lib="Homo sapiens NEUROBLASTOMA"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT 227 a 498 c 128 g 219 t 129 others
ORIGIN
Query Match 3.2%; Score 50; DB 13; Length 1201;
Best Local Similarity 37.3%; Pred. No. 0.091;
Matches 179; Conservative 43; Mismatches 258; Indels 0; Gaps 0;

QY 15 CCATCCCTCTCGGCTCCGTTACCAAGTCTGCTCAAGGATCGGCGCTCTTCTCTGTC 74
Db 244 CSCTMMACGCGCCCATCMCCCMCCCGCTCCYAGSCCTAACTCGCAYCCCCACCCCYC 303

QY 75 TCCCATCTCGAGCATGCCAATGCTCAGTACAGCATGAGCTCAGTACCCAGTCTCTCA 134
Db 304 ACCAAAMCTYACCTTGGCCATCCCTTTACSCCCCCCYCMCCCTCTCCMAAACCTCCC 363

QY 135 GCCGCACACCTGCTCTCTGCAACCCCATATAGCCAGTGCCTCCCTCCTCACTCGTA 194
Db 364 CCTCATCCCCCGCTCTCTCCCCCGCGYCCCCCCCCCTACCCCCCMWACMSCCCYCCCM 423

QY 195 CATGGGGAGCGCGGTACCGGCTGTATCTGAACAGGTACCCCGCATCAAGTCAAGTGT 254
Db 424 AGCCCGCGCGCGCTTCCACCCCCCAGCCCCCGCGCCCCCMCCCGCCSYCCCMCCCA 483
```

```
QY 255 TTACGGCTCTTCTGCTCGCGCATATATGCCCCACACTACCGTGGGAGCTTGCCTCCGAC 314
Db 484 TACMYCCCCCCCCCCCCSYGCCCCCYCCCCGGGCTCMCCCCCCCCSSSSCCCCCCC 543

QY 315 ATCTTTCTTCTCATCCCAATCGGAGGCGAGGCACAGGCGAGCAATCGCGCACTA 374
Db 544 AYCCCCCCCCCCCCCCCCCCCCCCCCCMCCMCCYCCCCCCCCCCCCCYCCCCCGT 603

QY 375 TCCTCCTCTCATAGCGTGTCTCCGCGCGCTTCCAGCGCTCAGTCTGATCCCGCAGCAAT 434
Db 604 CCCTCCCGCGCGGCGSSCCCCCCCCCGCGCCSSCCMCCCTMYSCCGCGCCSSCCTC 663

QY 435 TCGCGCGGCGCTCCCGGGACCGTCTGCTGACTTCAACAATGGACTTCTTTCAGGAGC 494
Db 664 CCCCCGGTCCCTCCACCCACCCACAGTATTAACAGTCTTAATGGCAATCACTGATAGGC 723

RESULT 13
BX421672
LOCUS
DEFINITION BX421672 Homo sapiens T CELLS (JURKAT CELL LINE) Homo sapiens cDNA
clone CS0DH003YC20 5-PRIME, mRNA sequence.
ACCESSION BX421672
VERSION BX421672.1 GI:30764106
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 1039.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DH003BB10QP1&cluster=1039.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DH003BB10QP1.
FEATURES
source
Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DH003YC20"
/tissue_type="T CELLS (JURKAT CELL LINE)"
/clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE)"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT 156 a 536 c 203 g 216 t 90 others
ORIGIN
Query Match 3.2%; Score 49.6; DB 13; Length 1201;
Best Local Similarity 38.8%; Pred. No. 0.12;
Matches 184; Conservative 33; Mismatches 257; Indels 0; Gaps 0;

QY 14 ACCATCCCTCTCGGCTCCGTTACCAAGTCTGCTCAAGGATCGGCGCTTCTCTCTG 73
Db 133 MCGMCCCMYIYTCMCCCGCTGMCCCTCTCCCTCCCMCAAGCCCCCCCCCCCCCTCC 192

QY 74 CTCCCATCTCGAGCATGCCAATGCTCAGTACAGTACAGTACAGTACCTCAGTACCTCTCTC 133
```


[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 2, 2004, 18:04:01 ; Search time 468 Seconds
(without alignments)
8980.823 Million cell updates/sec

Title: US-10-029-180-7

Perfect score: 1557

Sequence: 1 atggatcctagaaacatcc.....ccgttcagccgcagacaa 1557

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq 19Jun03.*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
- 25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1557	100.0	1557	24	ABQ94220 FLO11 gene expres
2	1557	100.0	1800	24	ABQ94276 Fungal gene expres
3	1557	100.0	1833	24	ABQ94275 Fungal gene expres
4	391.6	25.2	867	25	ABZ51240 Aspergillus oryzae
5	280.8	18.0	612	25	ABZ54133 Aspergillus oryzae
6	59	3.8	236	21	AAC57129 Pinus radiata tran
7	49.2	3.2	1856	23	ABL28791 Drosophila melanog
8	49.2	3.2	3219	23	ABL28790 Drosophila melanog

C	9	49.2	3.2	6530	14	AAQ51557	Loricrin gene. Ho
C	10	49.2	3.2	6530	20	AAZ22072	Nucleotide sequenc
C	11	49.2	3.2	6530	22	AAC68948	Mouse loricrin gen
	12	48.6	3.1	3073	25	AAD51683	Human nucleic acid
	13	46.6	3.0	549	21	AAC93730	Cat flea hindgut a
	14	46	3.0	4403765	22	AAI99683	Mycobacterium tube
	15	46	3.0	4411529	22	AAI99682	Mycobacterium tube
	16	45.2	2.9	492	17	AAT08579	Partial genomic DN
C	17	45.2	2.9	720	22	AAS05643	Mammalian vestibul
	18	45.2	2.9	1476	17	AAT08578	Zinc finger protei
	19	44.6	2.9	3269	23	ABL26283	Drosophila melanog
C	20	44.6	2.9	12920	23	ABL26282	Drosophila melanog
	21	44.4	2.9	4409	23	ABL17895	Drosophila melanog
	22	44.4	2.9	5021	23	ABL11783	Drosophila melanog
	23	44.4	2.9	7394	23	ABL17894	Drosophila melanog
	24	44.4	2.9	7614	23	ABL11782	Drosophila melanog
C	25	44.4	2.9	10732	21	AAA10594	Gene encoding a su
C	26	44.4	2.9	31814	25	AAD47150	Human Ras-like pro
C	27	43.8	2.8	2480	23	ABL25810	Drosophila melanog
	28	43.8	2.8	2874	23	ABL28677	Drosophila melanog
C	29	43.8	2.8	5125	23	ABL28676	Drosophila melanog
	30	43.2	2.8	646	22	AAS26107	Human CDNA encodin
	31	43.2	2.8	646	25	ABX73448	Human novel polynu
	32	43	2.8	336	20	AAV86698	EST clone BG228.
	33	43	2.8	1506	23	ABL29105	Drosophila melanog
	34	43	2.8	2336	21	AAZ35052	Mouse CNREB-2S (sh
	35	43	2.8	2651	21	AAZ35053	Mouse CNREB-2L (lo
	36	43	2.8	3215	23	ABL29108	Drosophila melanog
	37	43	2.8	3506	23	ABL29104	Drosophila melanog
	38	43	2.8	3847	21	AAZ35044	Mouse CNREB-2 cDNA
	39	43	2.8	4841	22	AAK52955	Human polynucleoti
	40	43	2.8	4880	22	AAK51971	Human polynucleoti
	41	43	2.8	4898	25	ABZ79896	Human nucleic-acid
C	42	43	2.8	16080	24	AAD28651	Human Sal2 gene.
	43	42.6	2.7	1134	23	ABL25805	Drosophila melanog
	44	42.6	2.7	1152	23	ABL13965	Drosophila melanog
	45	42.6	2.7	3213	23	ABL13964	Drosophila melanog

ALIGNMENTS

RESULT 1

ABQ94220
ID ABQ94220 standard; DNA; 1557 BP.

XX ABQ94220;

XX 22-OCT-2002 (first entry)

DT FLO11 gene expression regulator An09 coding sequence.

XX Fungal gene expression regulator; fungicide; gene therapy; An01; An05;
KW An09; An10; An13; An17; An20; An28; An34; At01-1; At01-2; At03; At05;
KW At07; At08; At11; At14; At16; At18; At19; At20; At22; At24; At27; At32;
KW Pc05; Pc06; Pc07; Pc08; Pc09; Pc10; Pc18; Pc23; Pc24; Pc25; Pc33; Pc34;
KW FLO11; fungal invasion; secondary metabolite; At279; At286; At291; At320;
KW At322; An1000; At167; At221; At233; At239; At240; At274; Pc1000; Pc1001;
KW lovF; lovE; lovastatin; Pc804; acvA; penicillin; antifungal; gene; ds.

XX Unidentified.

XX WO200257456-A2.

XX 25-JUL-2002.

XX 24-DEC-2001; 2001WO-US49911.

XX 22-DEC-2000; 2000US-257431P.

XX (MICR-) MICROBIA.

XX Cali BM, Madden KT, Milne TG, Zhang L, Silva JC, Trueheart J;

ID XX ABQ94276 standard; DNA; 1800 BP.
AC XX ABQ94276;
XX XX
DT 22-OCT-2002 (first entry)
XX XX
DE Fungal gene expression regulator coding sequence SEQ ID 118.
XX XX
KW Fungal gene expression regulator; fungicide; gene therapy; An01; An05;
KW An09; An10; An13; An17; An20; An28; An34; At01-1; At01-2; At03; At05;
KW At07; At08; At11; At14; At16; At18; At19; At20; At22; At24; At27; At32;
KW Pc05; Pc06; Pc07; Pc08; Pc09; Pc10; Pc18; Pc23; Pc24; Pc25; Pc33; Pc34;
KW FLO11; fungal invasion; secondary metabolite; At279; At286; At291; At320;
KW At322; An1000; At167; At221; At233; At239; At240; At274; Pc1000; Pc1001;
KW lovF; lovE; lovastatin; Pc804; acvA; penicillin; antifungal; gene; ds.
XX XX
OS Unidentified.
XX XX
PN WO200257456-A2.
XX XX
PD 25-JUL-2002.
XX XX
PF 24-DEC-2001; 2001WO-US49911.
XX XX
PR 22-DEC-2000; 2000US-257431P.
XX XX
PA (MICR-) MICROBIA.
XX XX
PI Cali BM, Madden KT, Milne TG, Zhang L, Silva JC, Trueheart J;
PI Holtzman D, Sherman A;
XX XX
XX WPI; 2002-627368/67.
XX XX
PT New isolated or recombinant gene, or purified protein, useful in
PT regulating fungal gene expression of FLO11, lovF, lovE or acvA for the
PT production of enzymes, secondary metabolites or other commercially and
PT medically useful products -
XX XX
PS Claim 67; Page 66-67; 71pp; English.
XX XX
CC The present invention relates to novel fungal gene expression regulators
CC (ABQ94217-ABQ94285 and ABP63084-ABP63152). An01, An05, An09, An10, An13,
CC An17, An20, An28, An34, At01-1, At01-2, At03, At05, At07, At08, At11,
CC At14, At16, At18, At19, At20, At22, At24, At27, At32, Pc05, Pc06, Pc07,
CC Pc08, Pc09, Pc10, Pc18, Pc23, Pc24, Pc25, Pc33 and Pc34 are FLO11 gene
CC expression regulators. FLO11 is required for fungal invasion and its
CC expression is believed to be regulated by factors that also modulate
CC secondary metabolite production. At279, At286, At291, At320, At322,
CC An1000, At167, At221, At233, At239, At240, At274, Pc1000 and Pc1001 are
CC lovF gene expression regulators, and At501 and At574 are lovE gene
CC expression regulators. lovF and lovE are believed to be involved in the
CC production of the secondary metabolite lovastatin. Pc804 is an acvA gene
CC expression regulator. acvA is involved in the production of the secondary
CC metabolite penicillin. The fungal gene expression regulators and their
CC coding sequences are useful in regulating or manipulating the expression
CC of fungal genes that are involved in the production of enzymes, secondary
CC metabolites and other commercially and medically useful products, in
CC order to achieve maximum benefit. The genes may also be used to identify
CC genes relevant to fungal invasion which may act as targets for the
CC development of antifungal drugs.
XX XX
SQ Sequence 1800 BP; 372 A; 606 C; 463 G; 359 T; 0 other;
Query Match 100.0%; Score 1557; DB 24; Length 1800;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGATCTTAGAAACCATCCCTCTCGGCTCCGTCACAGTCTGCCTCAAGGATCGGCG 60
DB 244 ATGGATCTTAGAAACCATCCCTCTCGGCTCCGTCACAGTCTGCCTCAAGGATCGGCG 303
QY 61 CCTCTTCTTCTGCTCCCATCTCGAGCATGCCAATGCCTCAGTACACGATCGAGCTCCAG 120
DB 1384 CCTGAAGACGCGTTTCCCGGCGCGCGCGCGCATCAAGGAGCCCTTGGGCCACAAGCTCCC 1443

Db 304 CCTCTTCTTCTGCTCCCATCTCGAGCATGCCAATGCCTCAGTACACGATGCAGCCTCAG 363
QY 121 TACCCAGTCTCTCAGCCGCACACACCTGCTCCTCTCTGCAACCCCATCATAGCCAGTCGCCC 180
Db 364 TACCCAGTCTCTCAGCCGCACACACCTGCTCCTCTCTGCAACCCCATCATAGCCAGTCGCCC 423
QY 181 GCTCCTCACTCGTACATGGGGCAGCCGCTACCGGCTGTATCTGAACAGGTACCCCGCA 240
Db 424 GCTCCTCACTCGTACATGGGGCAGCCGCTACCGGCTGTATCTGAACAGGTACCCCGCA 483
QY 241 TCAAGTACAGATGTTTACGCGTCTTCTGCTGCGCCGATATGCCCCACACTACCGTGGGC 300
Db 484 TCAAGTACAGATGTTTACGCGTCTTCTGCTGCGCCGATATGCCCCACACTACCGTGGGC 543
QY 301 AGCTTGCCTCCGACATCTTCTCTCTCATCCCAATCCGAGGCGCAGGCAGGCGCAG 360
Db 544 AGCTTGCCTCCGACATCTTCTCTCTCATCCCAATCCGAGGCGCAGGCAGGCGCAG 603
QY 361 CAATCGCCGCACTATCTCTCTCATAGCGTGTCTCCGCGCTTCCAGCGCTCAGTCG 420
Db 604 CAATCGCCGCACTATCTCTCTCATAGCGTGTCTCCGCGCTTCCAGCGCTCAGTCG 663
QY 421 TACCGCAGCCAATTTGCGCGCGCTCTCCCGGACCGTCTGCTGCTGACTTCAACAATGGA 480
Db 664 TACCGCAGCCAATTTGCGCGCGCTCTCCCGGACCGTCTGCTGCTGACTTCAACAATGGA 723
QY 481 CTTCCTTCAGGAGCATTCAGTTATTTCGACGGAAGGCTCAAGGTTGGACCCCGTTGCT 540
Db 724 CTTCCTTCAGGAGCATTCAGTTATTTCGACGGAAGGCTCAAGGTTGGACCCCGTTGCT 783
QY 541 GCGAATGGTGTCTGCGCGCTATCCCGGGAAGGACTCCCCCGGAACCCAGGTTGTTGTTCT 600
Db 784 GCGAATGGTGTCTGCGCGCTATCCCGGGAAGGACTCCCCCGGAACCCAGGTTGTTGTTCT 843
QY 601 CAGGGCGACGCGGTATCTTCCGAGTGTTCGCGGACGCGCAACTCCGGTCACAAATGGT 660
Db 844 CAGGGCGACGCGGTATCTTCCGAGTGTTCGCGGACGCGCAACTCCGGTCACAAATGGT 903
QY 661 GTTAACGGCACCGGCAAGAACACTATATCCCGGCAAGGATGCCGATGGAAAGTTCCCT 720
Db 904 GTTAACGGCACCGGCAAGAACACTATATCCCGGCAAGGATGCCGATGGAAAGTTCCCT 963
QY 721 TGCCGGAATGTAAACAGACTTATCTTCATGCCAAGCATCTCAAGCGCCATCTGCTACGC 780
Db 964 TGCCGGAATGTAAACAGACTTATCTTCATGCCAAGCATCTCAAGCGCCATCTGCTACGC 1023
QY 781 CACACTGGTGACCGCCGCTACATGTGTCTTTTGCAGAACACCTTCTCTCGCAGTGAT 840
Db 1024 CACACTGGTGACCGCCGCTACATGTGTCTTTTGCAGAACACCTTCTCTCGCAGTGAT 1083
QY 841 ATCTTGAAACGTCATTTCCAAATAATGCTCAATCAGGCGTGGTAACCCACCGGAGCAACG 900
Db 1084 ATCTTGAAACGTCATTTCCAAATAATGCTCAATCAGGCGTGGTAACCCACCGGAGCAACG 1143
QY 901 CACTTGTGCGACCCCAATGCGCATGTGAAGAGTCCCAACAGCAGGCTGCGGCGAATCCT 960
Db 1144 CACTTGTGCGACCCCAATGCGCATGTGAAGAGTCCCAACAGCAGGCTGCGGCGAATCCT 1203
QY 961 GTAAACCTGTCCAGGATGAAGTACAGTAGTACCGTCCCGCTTCCCAATGGCATCCCGGC 1020
Db 1204 GTAAACCTGTCCAGGATGAAGTACAGTAGTACCGTCCCGCTTCCCAATGGCATCCCGGC 1263
QY 1021 ACGACTTACGGCGAGGAGCGCTCAACGGCAATGGACTAGTCTCCGCGCGCGGCGGATAC 1080
Db 1264 ACGACTTACGGCGAGGAGCGCTCAACGGCAATGGACTAGTCTCCGCGCGCGGCGGATAC 1323
QY 1081 GCGGATCACCAGACTATGGGCTTCCCAATGTCTATCCGTCAACGGGATGGGCGGCTGCTCAG 1140
Db 1324 GCGGATCACCAGACTATGGGCTTCCCAATGTCTATCCGTCAACGGGATGGGCGGCTGCTCAG 1383
QY 1141 CCTGAAGACGCGTTTCCCGGCGCGCGCGCATCAAGGAGCCCTTGGGCCACAAGCTCCC 1200
Db 1384 CCTGAAGACGCGTTTCCCGGCGCGCGCGCATCAAGGAGCCCTTGGGCCACAAGCTCCC 1443

QY 1201 AAGCAGAGCCCGTATCTCGTGCAGCCGGGTGCTGACCCCTTCTGGCCACCAGTTGAATATT 1260
DB 1444 AAGCAGAGCCCGTATCTCGTGCAGCCGGGTGCTGACCCCTTCTGGCCACCAGTTGAATATT 1503
QY 1261 GACCGAAACATCGAGCAGGTAAACAAACCGGTTGTTCAAGACCCCAAGCCCTGTGATG 1320
DB 1504 GACCGAAACATCGAGCAGGTAAACAAACCGGTTGTTCAAGACCCCAAGCCCTGTGATG 1563
QY 1321 CCAGGACATCCCGGCCACCCCGGTGAGCTTGACTGGACGTCTATGTTCCAACTCAAGCT 1380
DB 1564 CCAGGACATCCCGGCCACCCCGGTGAGCTTGACTGGACGTCTATGTTCCAACTCAAGCT 1623
QY 1381 CCCGAGGGCTACATGTTCTCCAGTCTATGCCTGGTCAAGAGCCCATCCACGCTCAT 1440
DB 1624 CCCGAGGGCTACATGTTCTCCAGTCTATGCCTGGTCAAGAGCCCATCCACGCTCAT 1683
QY 1441 GTCGAGACCGAGCGAAAGTATTACCCCAACCACTACCGCTGGTCAAGAGAGTGAATGAAC 1500
DB 1684 GTCGAGACCGAGCGAAAGTATTACCCCAACCACTACCGCTGGTCAAGAGAGTGAATGAAC 1743
QY 1501 GGTCTCTATCTGGCTTCGACTATGAGTGGCGACGCCCGTTCAGCCCGCCAGACAA 1557
DB 1744 GGTCTCTATCTGGCTTCGACTATGAGTGGCGACGCCCGTTCAGCCCGCCAGACAA 1800

RESULT 3
ABQ94275

ABQ94275 standard; DNA; 1833 BP.

ABQ94275;

22-OCT-2002 (first entry)

Fungal gene expression regulator coding sequence SEQ ID 117.

Fungal gene expression regulator; fungicide; gene therapy; An01; An05;
An09; An10; An13; An17; An20; An28; An34; At01-1; At01-2; At03; At05;
At07; At08; At11; At14; At16; At18; At19; At20; At22; At24; At27; At32;
Pc05; Pc06; Pc07; Pc08; Pc09; Pc10; Pc18; Pc23; Pc24; Pc25; Pc33; Pc34;
FLO11; fungal invasion; secondary metabolite; At279; At286; At291; At320;
At322; An1000; At167; At221; At233; At239; At240; At274; Pc1000; Pc1001;
lovF; lovE; lovastatin; Pc804; acvA; penicillin; antifungal; gene; ds.

Unidentified.

WO200257456-A2.

25-JUL-2002.

24-DEC-2001; 2001WO-US49911.

22-DEC-2000; 2000US-257431P.

(MICR-) MICROBIA.

Cali BM, Madden KT, Milne TG, Zhang L, Silva JC, Trueheart J;
Holtzman D, Sherman A;

WPI; 2002-627368/67.

New isolated or recombinant gene, or purified protein, useful in
regulating fungal gene expression of FLO11, lovF, lovE or acvA for the
production of enzymes, secondary metabolites or other commercially and
medically useful products

Claim 67; Page 66; 71pp; English.

The present invention relates to novel fungal gene expression regulators
(ABQ94217-ABQ94285 and ABP63084-ABP63152). An01, An05, An09, An10, An13,
An17, An20, An28, An34, At01-1, At01-2, At03, At05, At07, At08, At11,
At14, At16, At18, At19, At20, At22, At24, At27, At32, Pc05, Pc06, Pc07,
Pc08, Pc09, Pc10, Pc18, Pc23, Pc24, Pc25, Pc33 and Pc34 are FLO11 gene

CC expression regulators. FLO11 is required for fungal invasion and its
expression is believed to be regulated by factors that also modulate
secondary metabolite production. At279, At286, At291, At320, At322,
An1000, At167, At221, At233, At239, At240, At274, Pc1000 and Pc1001 are
lovF gene expression regulators, and At501 and At574 are lovE gene
expression regulators. lovF and lovE are believed to be involved in the
production of the secondary metabolite lovastatin. Pc804 is an acvA gene
expression regulator. acvA is involved in the production of the secondary
metabolite penicillin. The fungal gene expression regulators and their
coding sequences are useful in regulating or manipulating the expression
of fungal genes that are involved in the production of enzymes, secondary
metabolites and other commercially and medically useful products, in
order to achieve maximum benefit. The genes may also be used to identify
genes relevant to fungal invasion which may act as targets for the
development of antifungal drugs.

XX
SQ Sequence 1833 BP; 386 A; 613 C; 469 G; 365 T; 0 other;

Query Match 100.0%; Score 1557; DB 24; Length 1833;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGATCCTAGAAACCAATCCCTCTCGGCCTCCGTCTACCAAGTCTCAAGGATCGGCG 60
DB 277 ATGGATCCTAGAAACCAATCCCTCTCGGCCTCCGTCTACCAAGTCTCAAGGATCGGCG 336
QY 61 CCTCTTCTCTGCTCCCATCTCGAGCATGCCAATGCCTCAGTACACGATGCAGCTCAG 120
DB 337 CCTCTTCTCTGCTCCCATCTCGAGCATGCCAATGCCTCAGTACACGATGCAGCTCAG 396
QY 121 TACCCAGTCTCTCAGCCCGCACACCTCTGCTCTCTCTGCAACCCCATCATAGCCAGTGGCC 180
DB 397 TACCCAGTCTCTCAGCCCGCACACCTCTGCTCTCTCTGCAACCCCATCATAGCCAGTGGCC 456
QY 181 GTCCTCACTCGTACATGGGGGACGCCCGCTACCGGCTGATCTGAACAGGTACCCCGCA 240
DB 457 GTCCTCACTCGTACATGGGGGACGCCCGCTACCGGCTGATCTGAACAGGTACCCCGCA 516
QY 241 TCAAGTCACGATGTTTACGCGTCTTCTGCTGCGCCGATATATGCCACACTACCGTGGC 300
DB 517 TCAAGTCACGATGTTTACGCGTCTTCTGCTGCGCCGATATATGCCACACTACCGTGGC 576
QY 301 AGCTTGCTCCGACATCTTCTCTCTCTCATCCCAATCCGAGGCGGACGACAGGCGCAG 360
DB 577 AGCTTGCTCCGACATCTTCTCTCTCTCATCCCAATCCGAGGCGGACGACAGGCGCAG 636
QY 361 CAATCGCCGCACTATCTCTCTCTCATAGCGTGTCTCCCGCCGCTTCCAGCGCTCAGTCG 420
DB 637 CAATCGCCGCACTATCTCTCTCTCATAGCGTGTCTCCCGCCGCTTCCAGCGCTCAGTCG 696
QY 421 TACCCGACGCAATTTGGCCGGCCCTTCCCGGGACCGTCTGCTGCTGACTTCAACAATGA 480
DB 697 TACCCGACGCAATTTGGCCGGCCCTTCCCGGGACCGTCTGCTGCTGACTTCAACAATGA 756
QY 481 CTTCTTTCAGGAGCATTGAGTTATTCGACGGAAGCCTCAAGGTTGGGACCCCGTTGCT 540
DB 757 CTTCTTTCAGGAGCATTGAGTTATTCGACGGAAGCCTCAAGGTTGGGACCCCGTTGCT 816
QY 541 GCGAATGCTGCTGCGCGTATCCCGGGAAGGACTCCCCCGAACCCAGGTTGTTGTTCT 600
DB 817 GCGAATGCTGCTGCGCGTATCCCGGGAAGGACTCCCCCGAACCCAGGTTGTTGTTCT 876
QY 601 CAGGGGCGACGCGGTATCTTCCGAGTGTTCGGGACGCGCAACTCCGGTCAACAATGGT 660
DB 877 CAGGGGCGACGCGGTATCTTCCGAGTGTTCGGGACGCGCAACTCCGGTCAACAATGGT 936
QY 661 GTTAACGGCACCGGCAAGACACTACTATCTCCCGCCCAAGGATGCCGATGGAAGTTCCCT 720
DB 937 GTTAACGGCACCGGCAAGACACTACTATCTCCCGCCCAAGGATGCCGATGGAAGTTCCCT 996
QY 721 TGCCCGGAAGTGAACAAGACTTATCTTCAAGCAAGCATCTCAAGCGGCCACTGTCTACGC 780
DB 997 TGCCCGGAAGTGAACAAGACTTATCTTCAAGCAAGCATCTCAAGCGGCCACTGTCTACGC 1056

QY 781 CACACTGGTGACCGCCGCTACATGTGTGTTCTTTGCAAGACACCTTCTCTGCGAGTGAT 840
Db |||||
QY 1057 CACACTGGTGACCGCCGCTACATGTGTGTTCTTTGCAAGACACCTTCTCTGCGAGTGAT 1116
Db |||||
QY 841 ATCCTGAACCGTCAATTTCCAAAAATGCTCAATCAGCGTGGTAAACCCACCGGAGCAACG 900
Db |||||
QY 1117 ATCCTGAACCGTCAATTTCCAAAAATGCTCAATCAGCGTGGTAAACCCACCGGAGCAACG 1176
QY 901 CACTTGTGCGACCCCAATGCGCATGTGAAGAGGTCCCAACAGCAGGCTGCGGGAATCCT 960
Db |||||
QY 1177 CACTTGTGCGACCCCAATGCGCATGTGAAGAGGTCCCAACAGCAGGCTGCGGGAATCCT 1236
QY 961 GTAAACCTGTCCAGGATGAAGTCAAGTACCGTCCCGCTCCCAATGGCATCCCGGC 1020
Db |||||
QY 1237 GTAAACCTGTCCAGGATGAAGTCAAGTACCGTCCCGCTCCCAATGGCATCCCGGC 1296
QY 1021 ACGACTTACGGGAGGAGCGCTCAACGGCAATGGAATGACTAGTCTCCGCCCGGCGGAGGTAC 1080
Db |||||
QY 1297 ACGACTTACGGGAGGAGCGCTCAACGGCAATGGAATGACTAGTCTCCGCCCGGCGGAGGTAC 1356
QY 1081 GCGGATCACAGACTATGGGCTTCCCAATGTATCGGTCAACGGGATGGCGGTGTCAG 1140
Db |||||
QY 1357 GCGGATCACAGACTATGGGCTTCCCAATGTATCGGTCAACGGGATGGCGGTGTCAG 1416
QY 1141 CCTGAAGACGCGTTTCCGGCGGCGCGCATCAAGGAGCCCTTGGCCACAAGCTCCC 1200
Db |||||
QY 1417 CCTGAAGACGCGTTTCCGGCGGCGCGCATCAAGGAGCCCTTGGCCACAAGCTCCC 1476
QY 1201 AAGCAGAGCCCGTATCTGTCGAGCCGGGTGCTGACCCCTTCTGGCCACCAATGATTT 1260
Db |||||
QY 1477 AAGCAGAGCCCGTATCTGTCGAGCCGGGTGCTGACCCCTTCTGGCCACCAATGATTT 1536
QY 1261 GACCGAACAATCGAGAGGTAAACAAACCGGTTGTTCAAGACCCCAAGCGCCCTGTGATG 1320
Db |||||
QY 1537 GACCGAACAATCGAGAGGTAAACAAACCGGTTGTTCAAGACCCCAAGCGCCCTGTGATG 1596
QY 1321 CCAGGACATCCCGGCAACCCCGGTGAGCTTGACTGGACGTCTATGTTCCAACTCAAGCT 1380
Db |||||
QY 1597 CCAGGACATCCCGGCAACCCCGGTGAGCTTGACTGGACGTCTATGTTCCAACTCAAGCT 1656
QY 1381 CCCGAGGCTACATGTTCTCCAGTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
Db |||||
QY 1657 CCCGAGGCTACATGTTCTCCAGTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1716
QY 1441 GTCGAGACCGAGCGAAAGTATTACCCCACTACCGTGGTCAAGAGAGTGAATGAAC 1500
Db |||||
QY 1717 GTCGAGACCGAGCGAAAGTATTACCCCACTACCGTGGTCAAGAGAGTGAATGAAC 1776
QY 1501 GGTCTCTATCTGCTTCCGACTATGAGTGGCGGCGGACCGTTTCAGCCCGCCAGACAA 1557
Db |||||
QY 1777 GGTCTCTATCTGCTTCCGACTATGAGTGGCGGCGGACCGTTTCAGCCCGCCAGACAA 1833

RESULT 4
ABZ51240
ID ABZ51240 standard; cDNA; 867 BP.

AC ABZ51240;
XX
XX
DT 28-MAR-2003 (first entry)
XX
DE Aspergillus oryzae polynucleotide SEQ ID NO 353.
XX
XX
KW Aspergillus oryzae; fermentation; fungus; industrial; EST;
KW expressed sequence tag; gene; ss.
XX
OS Aspergillus oryzae.
XX
PN WO200279476-A1.
XX
PD 10-OCT-2002.
XX

PF 22-MAR-2002; 2002WO-IB00890.
XX
PR 30-MAR-2001; 2001JP-0098371.
XX
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (NARE-) NAT RES INST BREWING.
PA (NORQ) NAT FOOD RES INST MIN AGRIC.
XX
PI Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;
PI Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;
XX
DR WPI; 2003-046817/04.
XX
PT Detection of expression of specific Aspergillus genes for monitoring
PT the fermentation and growth conditions of the fungus, using DNA probes
PT
XX
PS Claim 1; SEQ ID NO 353; 48pp + Sequence Listing; Japanese.
XX
CC The invention relates to a polynucleotide having any of 6006 specific
CC sequences (ABZ50888-ABZ56893), which are expressed by a fungus under
CC specific culture conditions including one or more of eutrophic,
CC oligotrophic, solid, early germination, alkaline, high temperature, low
CC temperature or maltose culture or polynucleotides stringently hybridising
CC to these sequences. The polynucleotides are useful for monitoring the
CC progress of fermentation and the growth conditions of a fungus,
CC especially of Aspergillus oryzae which is widely used in industrial
CC fermentation. Also monitoring for fungal contamination.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 867 BP; 188 A; 257 C; 234 G; 182 T; 6 other;
Query Match 25.2%; Score 391.6; DB 25; Length 867;
Best Local Similarity 71.6%; Pred. No. 1.6e-103;
Matches 591; Conservative 1; Mismatches 206; Indels 27; Gaps 5;

QY 751 GCCAAGCATCTCAAGCGCCATCTCTCGAGTGATATCTGAAACGTCATTTCCAAAAATGCTCA 810
Db |||||
QY 811 CTTTGCAAGACACCTTCTCTCGAGTGATATCTGAAACGTCATTTCCAAAAATGCTCA 870
Db |||||
QY 91 CTGTGCAAGGATATCTTCTCCGTAAGTATCTGAAAGCGCCACTTCCAGAGTGTTCG 150
QY 871 ATCAGCGTGGTAAACCCACCGGAGCAACGCACTTGTGCAACCCCAATGCGCATGTGAAG 930
Db |||||
QY 151 TTACGACGCGGCAACCCGACGCGGAGCGACCCACTTGTGCAATCCCGAGGCACATTTGAAG 210
QY 931 AGGTCCCAACAGCAGGCTGCGGCGAATCTGTAAACCTGTCCAGGATGAAGTCAGTAGT 990
Db |||||
QY 211 AGGTCTCAA-----GCTGGAATCTGCTAAACCGGTTAGGATGAAGTCAGTAGT 261
QY 991 ACCGTCCCGCTCCCAATGGCATCCCGGCGACGACTTACGGCGAGGAGCGGTCACCGGC 1050
Db |||||
QY 262 ACCGTCTCTCCCGCACTGCCCTTCCGGGTACGGCTTACGGCGAGGGGCGCGTGAACGGT 321
QY 1051 AATGACTAGTCTCCGGCCCGGCGGAGGTACGCGGATCACCAAGTATGGGCTTCCCAATG 1110
Db |||||
QY 322 AACGGGCTGGCTTCAGGCGGACCTGGGTTACCGGATCAGCAGCCTCTGGGCTATCCGATG 381
QY 1111 TCATCCGTCAACGGGATGGCGGTGTCAGCTGAGACGCGCTTCCCGGCGGCGCGCG 1170
Db |||||
QY 382 TCGTCGGTCAATGGGATGAACCGTGGTCAACCGGACGATGCGTTCCTCCGCTGGTCAGCCA 441
QY 1171 CATCAAGGAGCCCTTGGCCACAAGCTCCCAAGCAGAGCCCGTATCTCGTCAGCCGGGT 1230
Db |||||
QY 442 CATCAGAGAGGCCCTTGGCTGGCGGCTCCCAAGCAGAACCCGATCTCTGGCGCAACCTGGC 501
QY 1231 GCTGACCCCTTCTGGCCACCAAGTGAATATTGACCGAAACATCGAGCAGGTAAACACCG 1290
Db |||||
QY 502 ACTGATGTATCTCAGCAGCTGAGTGTGTGACCGCTCTTCTTATGAACAGGTAAAGCCCGCG 561

XX 25-NOV-1999 (first entry)
DT Nucleotide sequence of the loricrin gene.

XX Vector; epidermal cell; loricrin gene; K6 keratin gene; wound;
KW surgical incision; disease; psoriasis; cancer; alopecia;
KW transgenic animal; ss.

XX Mus sp.

XX US5958764-A.

XX 28-SEP-1999.

XX 01-NOV-1993; 93US-0146930.

XX 30-APR-1992; 92US-0876286.

XX 29-OCT-1993; 93US-0145388.

XX (BAYU) BAYLOR COLLEGE MEDICINE.

XX Greenhalgh DA, Rothnagel JA, Roop DR;

XX WPI; 1999-561044/47.

XX Gene therapy vectors for expression in the epidermis -

XX Example 1; Columns 25-32; 34pp; English.

XX The specification describes a vector for the expression of a nucleic acid sequence in an epidermal cell, where the nucleic acid sequence is not a mammalian loricrin gene or a mammalian K6 keratin gene. The vector comprises a 5' flanking region including the sequences required for its expression; a 3' region from a mammalian loricrin or a mammalian K6 keratin gene which regulates expression (predominantly in an epidermal cell) of the nucleic acid sequence with which it is associated; and a linker connecting the 5' flanking region to the nucleic acid sequence, where the linker comprises a position for inserting the nucleic acid sequence and lacks the coding sequence of a gene with which it is naturally associated. The vector is able to encode a hormone, a growth factor, an enzyme, a clotting factor, an apolipoprotein, a receptor, a drug or an antigen. The vectors can be used to treat wounds or surgical incisions as well as diseases, such as psoriasis, cancer, alopecia and to create transgenic animals for assessing human disease in an animal model. The present sequence represents the loricrin gene.

XX Sequence 6530 BP; 1635 A; 1507 C; 1696 G; 1692 T; 0 other;

Query Match 3.2%; Score 49.2; DB 20; Length 6530;
Best Local Similarity 46.1%; Pred. No. 0.002;
Matches 165; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

QY 96 GCCTCAGTACAGGATGCGCTCAGTACCCAGTCTCTCAGCCGCGACACCTGCCTCCTCT 155

DB 3530 GCCTCCGTAGCTCTGGCACTGATCTGTTGGACGATCCGCCGCCGCCACCGGAGGA 3471

QY 156 GCAACCCCATATAGCCAGTCGCCCGCTCCTCACTCGTACATGGGGCAGCCCGTACCG 215

DB 3470 ACCACCTCGCAGCTCGTGC CGCGCCGACCGGAATAGCCGCCGCCGCCGCGCAGCTGGA 3411

QY 216 GCCTGATCTGAACAGGTACCCCGCATCAAGTCACGATGTTTACGGCTCTTCTGTCGCC 275

DB 3410 TCCACCGCACCTGAGTAGCCCGCCGACAGCTAGACCGCGCTGCTTCCACCGCGCC 3351

QY 276 GATAATGCCCCACACTACCGTGGGCGCTTGCCTCCGACATCTTCTCTCATCCCAA 335

DB 3350 GGAATAGCGCCTCCGACGCTGGAGCCACCGCCGCCACAGAGTACTTGCCTCCGGA 3291

QY 336 TCCGCGCGCGCAGGCACAGGGCGAGCAATGCCGCACTATCTCTCTCTCATAGGTGCT 395

DB 3290 ACCGCCGCGCAGCCACCGGAGCTGCCGCTCCAGAGGAGCCTCCGCGAGTAGGAGCGCC 3231

QY 396 CCGCGCCGCTTCCAGCGCTCAGTCGTACCCGAGCCAAATTGCGCGCGCTCCCGG 453
DB 3230 GCCTCCTGATCCGCTCCGCGAGCTGGAGCCGCCGCCGCCGCCGCGGAGGAGCCGCG 3173

RESULT 11

AAC68948/c

ID AAC68948 standard; DNA; 6530 BP.

XX AAC68948;

XX 26-FEB-2001 (first entry)

XX Mouse loricrin gene.

XX Loricrin; cytostatic; gene therapy; cell proliferation inhibition;
KW murine; skin cancer; skin ulcer; psoriasis; alopecia; ds.

XX Mus sp.

XX US6143727-A.

XX 07-NOV-2000.

XX 05-JUN-1995; 95US-0458240.

XX 01-NOV-1993; 93US-0146930.

XX 30-APR-1992; 92US-0876286.

XX 29-OCT-1993; 93US-0145388.

XX (BAYU) BAYLOR COLLEGE MEDICINE.

XX Roop DR, Greenhalgh DA, Rothnagel JA;

XX WPI; 2001-006441/01.

XX Gene therapy of skin cancer involving administering specific expression vector, loricrin or K6 keratin expression vector within the epidermis or epidermal cells

XX Claim 8; Columns 25-32; 35pp; English.

XX The present invention relates to a method for treating skin cancer. The method comprises administering a loricrin or K6 keratin expression vector comprising a nucleic acid sequence encoding a wild-type p53 protein at or directly around the site of skin cancer cell, where the expression of the p53 protein by the skin cancer cell results in inhibition of the cell's proliferation. The present sequence is the murine loricrin gene. This gene was used to construct the loricrin expression vector of the present invention. The method of the present invention may also be used to treat skin ulcers, psoriasis and alopecia.

XX Sequence 6530 BP; 1635 A; 1507 C; 1696 G; 1692 T; 0 other;

Query Match 3.2%; Score 49.2; DB 22; Length 6530;
Best Local Similarity 46.1%; Pred. No. 0.002;
Matches 165; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

QY 96 GCCTCAGTACAGGATGCGCTCAGTACCCAGTCTCTCAGCCGCGACACCTGCCTCCTCT 155

DB 3530 GCCTCCGTAGCTCTGGCACTGATCTGTTGGACGATCCGCCGCCGCCACCGGAGGA 3471

QY 156 GCAACCCCATATAGCCAGTCGCCCGCTCCTCACTCGTACATGGGGCAGCCCGTACCG 215

DB 3470 ACCACCTCGCAGCTCGTGC CGCGCCGACCGGAATAGCCGCCGCCGCCGCGCAGCTGGA 3411

QY 216 GCCTGATCTGAACAGGTACCCCGCATCAAGTCACGATGTTTACGGCTCTTCTGTCGCC 275

DB 3410 TCCACCGCACCTGAGTAGCCCGCCGACAGCTAGACCGCGCTGCTTCCACCGCGCC 3351

QY 276 GATAATGCCCCACACTACCGTGGGCGAGCTTGCCTCCGACATCTTCTCTCATCCCAA 335

DB 3350 GGAATAGCGCCTCCGACGCTGGAGCCACCGCCGCCACAGAGTACTTGCCTCCGGA 3291

QY 336 TCCGAGCGGCGAGGACAGCGGCGAGCAATCGCGCACTATCCTCCTCATAGCGTGCT 395
|||||
Db 3290 ACCGCGCGCGAGCCACCGAGCTGCCCTCCAGAGGAGCTCCGCGATGAGGCGGCC 3231
|||||
QY 396 CCCGCCGCTTCCAGCGCTCAGTCGTACCCGCGAGCAATGCGCGGCGCTCCCGG 453
|||||
Db 3230 GCCTCCTGATCCGCTCCGAGCTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 3173
|||||

RESULT 12
AAD51683
ID AAD51683 standard; cDNA; 3073 BP.
XX
AC AAD51683;
XX
DT 16-APR-2003 (first entry)
XX
DE Human nucleic acid associated protein (NAAP)-9 encoding cDNA.
XX
KW Human; nucleic acid associated protein; NAAP; cancer; atherosclerosis;
KW cell proliferative disorder; neurological disorder; Huntington's disease;
KW epilepsy; stroke; immune disorder; acquired immune deficiency syndrome;
KW AIDS; inflammatory disorder; allergy; developmental disorder; infection;
KW hypothyroidism; cushing's syndrome; gene therapy; cytostatic; nootropic;
KW anticonvulsant; neuroprotective; cerebroprotective; thyromimetic; gene;
KW ss.

XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT CDS 1055..2242
FT /*tag= a
FT /product= "Human NAAP-9 protein"
XX
PN WO200299115-A2.
XX
PD 12-DEC-2002.
XX
PF 31-MAY-2002; 2002WO-US17050.
XX
PR 01-JUN-2001; 2001US-295359P.
PR 08-JUN-2001; 2001US-296878P.
PR 08-JUN-2001; 2001US-297222P.
PR 15-JUN-2001; 2001US-298615P.
PR 15-JUN-2001; 2001US-298665P.
PR 15-JUN-2001; 2001US-298693P.
PR 21-JUN-2001; 2001US-300176P.
PR 19-APR-2002; 2002US-373891P.

XX (INCY-) INCYTE GENOMICS INC.
PA (YUEH/) YUE H.
XX
PI Yue H, Tang YT, Baughn MR, Becha SD, Warren BA, Walia NK, Lal PG;
PI Lee EA, Hafalia AJA, Richardson TW, Griffin JA, Emerling BM;
PI Ramkumar J, Yue H, Swarnakar A, Tran B, Li JX, Yao MG, Yang J;
PI Ison CH, Forsythe IJ, Honchell CD, Arvizu CS, Elliott VS, Lu Y;
PI Ding L, Luo W, Wang YE, Burford N, Borowsky ML, Nguyen DB;
PI Chinn AM, Kable AE;
XX
XX WPI; 2003-140626/13.
DR P-PSDB; AAE33769.
XX
XX New human nucleic acid associated proteins (NAAP), useful for
PT diagnosing, treating and preventing diseases or conditions associated
PT with the aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis,
PT epilepsy, or infections
XX
PS Claim 94; Column 235-236; 257pp; English.

XX
CC The present invention relates to human nucleic acid associated proteins
CC (NAAP) and polynucleotides encoding such proteins. NAAP sequences are
CC useful in diagnosing, treating and preventing diseases or conditions

CC associated with the decreased expression or overexpression of NAAP such
CC as cell proliferative (e.g. cancer, atherosclerosis), neurological (e.g.
CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,
CC allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)
CC disorders or infections. They are also used in gene therapy. The present
CC sequence is human NAAP-9 cDNA.

XX
SQ Sequence 3073 BP; 487 A; 1117 C; 970 G; 499 T; 0 other;
Query Match 3.1%; Score 48.6; DB 25; Length 3073;
Best Local Similarity 54.9%; Pred. No. 0.0021;
Matches 96; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 695 CCAAGGATGCCGATGGAAGTTCCCTTGCCCGCACTGTAAACAAGACTTATCTTCATGCCA 754
|||||
Db 86 CCGGGGTGGCGATGGCCCGTTCGCTGCCCACTCTGTGGAAGTTTCAAGAAGCCCCA 145
|||||
QY 755 AGCATCTCAAGCGCCATCTGTACGCCACACTGGTGACCGCCGACATGTGTCTTT 814
|||||
Db 146 GTCACCTCCACCGACAGATCATCCACACGGCGGAGAGCCCTTCTCCTGCTCGTGT 205
|||||
QY 815 GCAAAGACACCTTCTCGCAGTGATATCTGAAACGTCATTTCCAAAATGCTC 869
|||||
Db 206 GCAGCAAAAGCTTCAACCGCAGGAGAGTCTGAAGCGCCACGCTGAAGACGCACCTC 260
|||||

RESULT 13
AAC93730
ID AAC93730 standard; cDNA; 549 BP.
XX
AC AAC93730;
XX
DT 19-FEB-2001 (first entry)
XX
DE Cat flea hindgut and Malpighian tubule (HMT) cDNA, SEQ ID NO:225.
XX
KW Cat flea; hindgut and Malpighian tubule nucleic acid; HMT;
KW flea infestation; vaccine; antiparasitic; therapeutic target;
KW diagnosis; detection; ss.
XX
OS Ctenocephalides felis.
XX
PN WO200061621-A2.
XX
PD 19-OCT-2000.
XX
PF 07-APR-2000; 2000WO-US09437.
XX
PR 09-APR-1999; 99US-0128704.
XX
PA (HESK-) HESKA CORP.

XX
PI Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N;
XX
DR WPI; 2000-656323/63.
XX
PT Flea Malpighian tubule and head and nerve cord tissue derived nucleic
PT acids useful for the prevention, diagnosis and treatment of flea
PT infestations -
XX
PS Claim 26; Page 320; 964pp; English.
XX
CC The invention relates to novel cat flea (Ctenocephalides felis) nucleic
CC acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
CC or head and nerve cord (HNC) tissue. The invention also relates to the
CC encoded proteins. The invention additionally encompasses expression
CC constructs, recombinant viruses and recombinant cells comprising the
CC nucleic acids of the invention, recombinant production of the proteins,
CC antibodies against the proteins, a method of identifying inhibitors of
CC the proteins, and compositions comprising the inhibitors for
CC administration to an animal. The nucleic acids, and the proteins they
CC encode may be used in the prevention, treatment and diagnosis of diseases
CC associated with flea infestations. For example, the nucleic acids may be

XX (GENO-) INST GENOMIC RES.
XX
PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX
XX
DR WPI; 2001-647261/74.
XX
PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
PT 1551 and H37Rv differ -
XX
PS Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
XX
CC The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;

Query Match 3.0%; Score 46; DB 22; Length 4411529;
Best Local Similarity 45.8%; Pred. No. 0.31;
Matches 197; Conservative 0; Mismatches 230; Indels 3; Gaps 1;

QY 27 GCCTCCGTCTACAGTCTGCTCAAGATCGGCGCTCTTCTTCTGCTCCCATCTCGAG 86
Db 1631626 GCCGCGTTCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 1631685

QY 87 CATGCCAATGCCCTCAGTACAG---ATGCAGCCTCAGTACCCAGTCTCTCAGCGGCACAC 143
Db 1631686 GCCGCGTTCG 1631745

QY 144 CCTGCCTCTCTGCAACCCCATCATAGCCAGTCGCCCGCTCCTCACTCGTACATGGGCA 203
Db 1631746 CGCGCGTTCGCTGACTACGCTTGACCGCGCTTGCGCGCGCGCGCGCGCGCGCG 1631805

QY 204 GCCGCGGTACCGCGCTGATCTGAACAGGTACCCCGCATCAAGTCACGATGTTACGGTC 263
Db 1631806 GCCGCGTTCG 1631865

QY 264 TTCTGCTGCGCGGATATGCCCCACACTACCGTGGGCGAGCTTGCCTCCGACATCTTCCT 323
Db 1631866 ACCGCGCGCGCGCTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTT 1631925

QY 324 TTCTCATCCCAATCCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 383
Db 1631926 ACCGCGGTTCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGTTAC 1631985

QY 384 TCATAGCGTGTCTCCCGCGCGCTTCCAGCGCTCAGTCGTACCGCGAGCAATTGGCGCGGC 443
Db 1631986 GTTGGCGCGGTTTTCACAGGTTGGTGGCGTTGGCACTCATGCCACCAACCGCGCG 1632045

QY 444 GCCTCCCCCG 453.
Db 1632046 GCCGCGCGG 1632055

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 2, 2004, 18:05:06 ; Search time 5899 Seconds
(without alignments)
10797.811 Million cell updates/sec

Title: US-10-029-180-7
Perfect score: 1557
Sequence: 1 atggatcctagaacacatcc.....ccgttcagccgcagacaa 1557

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sta:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sta:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rod:*
- 36: em_htg_mam:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1557	100.0	1557	6	AX505141	AX505141 Sequence
2	1557	100.0	1560	8	AF467711	AF467711 Emericell
3	1557	100.0	1800	6	AX505252	AX505252 Sequence
4	1557	100.0	1833	6	AX505251	AX505251 Sequence
5	236.2	15.2	741	11	PM6H12G	AL686183 Penicilli
6	61	3.9	7218	6	I66494	I66494 Sequence 14
7	59.6	3.8	241859	2	AC136406	AC136406 Rattus no
8	57.2	3.7	214690	2	AC083889	AC083889 Mus muscu
9	57.2	3.7	219559	10	AC087329	AC087329 Mus Muscu
10	56.8	3.6	125020	9	AF429315	AF429315 Homo sapi
11	52.2	3.4	227533	2	AC094895	AC094895 Rattus no
12	52	3.3	178757	10	AC122509	AC122509 Mus muscu
13	51.8	3.3	197714	2	AC128967	AC128967 Rattus no
14	51.6	3.3	264103	2	AC126651	AC126651 Rattus no
15	51	3.3	5452	6	AX646997	AX646997 Sequence
16	51	3.3	298166	2	AC087563	AC087563 Homo sapi
17	50.8	3.3	236054	2	AC114393	AC114393 Rattus no
18	50	3.2	210582	2	BX324212	BX324212 Danio rer
19	50	3.2	232885	2	AC106163	AC106163 Rattus no
20	49.6	3.2	87120	2	AC012225	AC012225 Homo sapi
21	49.4	3.2	33058	2	AC100211	AC100211 Mus muscu
22	49.2	3.2	694	3	DMHSP70AB	Y14393 Drosophila
23	49.2	3.2	1751	10	MUSLRCNA	M34398 Mouse loric
24	49.2	3.2	1804	10	BC026781	BC026781 Mus muscu
25	49.2	3.2	1863	3	AY052029	AY052029 Drosophil
26	49.2	3.2	6530	6	AR076241	AR076241 Sequence
27	49.2	3.2	6530	10	MMU09189	U09189 Mus musculu
28	49.2	3.2	55528	3	AC105764	AC105764 Caenorhab
29	49.2	3.2	144152	3	AC007725	AC007725 Drosophil
30	49.2	3.2	182183	3	AC007889	AC007889 Drosophil
31	49.2	3.2	226782	3	AE003693	AE003693 Drosophil
32	49.2	3.2	255454	2	AC138676	AC138676 Mus muscu
33	48.8	3.1	209887	2	AC079424	AC079424 Mus muscu
34	48.6	3.1	156506	2	AC068038	AC068038 Homo sapi
35	48.6	3.1	194864	10	AL691489	AL691489 Mouse DNA
36	48.6	3.1	223879	9	AC008735	AC008735 Homo sapi
37	48.4	3.1	183991	2	BX072578	BX072578 Danio rer
38	48	3.1	5880	14	HSE4	D14486 Equine herp
39	47.4	3.0	188974	2	AC022682	AC022682 Mus muscu
40	47.4	3.0	303091	2	AC084799	AC084799 Mus muscu
41	47	3.0	161002	2	BX005475	BX005475 Danio rer
42	46.8	3.0	1349	3	AY122249	AY122249 Drosophil
43	46.8	3.0	85968	2	AC009536	AC009536 Drosophil
44	46.8	3.0	181773	2	AC136125	AC136125 Rattus no
45	46.8	3.0	212493	2	AC080167	AC080167 Mus muscu

ALIGNMENTS

RESULT 1	AX505141	AX505141	1557 bp	DNA	linear	PAT 27-SEP-2002
LOCUS	Sequence 7 from Patent WO02057456.					
DEFINITION	AX505141					
ACCESSION	AX505141					
VERSION	AX505141.1	GI:23386445				
KEYWORDS						
SOURCE	synthetic construct					
ORGANISM	synthetic construct					
REFERENCE	artificial sequences.					
AUTHORS	1					
TITLE	Cali,B.M., Madden,K.T., Milne,T.G., Zhang,L., Silva,J.C.,					
JOURNAL	Trueheart,J., Holtzman,D. and Sherman,A.					
	Regulators of fungal gene expression					
	Patent: WO 02057456-A 7 25-JUL-2002;					

Microbia (US)
Location/Qualifiers
1. 1557
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/note="fungal gene"
BASE COUNT 329 a 533 c 385 g 310 t
ORIGIN
Query Match 100.0%; Score 1557; DB 6; Length 1557;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 ATGGATCCTAGAAACCAATCCTCTCGGCCTCCGTCACAGTCTGCCTCAAGGATCGGCG 60
1 ATGGATCCTAGAAACCAATCCTCTCGGCCTCCGTCACAGTCTGCCTCAAGGATCGGCG 60
61 CCTCTTCTCTGCTCCCAATCTCGAGCATGCCAATGCCCTCAGTACAGGATCGGCTCAG 120
61 CCTCTTCTCTGCTCCCAATCTCGAGCATGCCAATGCCCTCAGTACAGGATCGGCTCAG 120
121 TACCCAGTCTCTCAGCCGACACCCCTGCTCCTCTGCAACCCCATCATAGCCAGTCCGCC 180
121 TACCCAGTCTCTCAGCCGACACCCCTGCTCCTCTGCAACCCCATCATAGCCAGTCCGCC 180
181 GCTCCTCACTCGTACATGGGGGAGCGCCGCTACCGGCTGATCTGAACAGGTACCCGCA 240
181 GCTCCTCACTCGTACATGGGGGAGCGCCGCTACCGGCTGATCTGAACAGGTACCCGCA 240
241 TCAAGTACGATGTTTACGCGTCTTCTGCTGCGCGGATATGCCCCACACTACCGTGGGC 300
241 TCAAGTACGATGTTTACGCGTCTTCTGCTGCGCGGATATGCCCCACACTACCGTGGGC 300
301 AGCTTGCCTCCGACATCTTCTCTCTCATAGCGTCTCCCGGACCGCTCAGTCAACAATGGA 360
301 AGCTTGCCTCCGACATCTTCTCTCTCATAGCGTCTCCCGGACCGCTCAGTCAACAATGGA 360
361 CAATCGCGCACTATCTCTCTCTCATAGCGTCTCCCGGACCGCTCAGTCAACAATGGA 420
361 CAATCGCGCACTATCTCTCTCTCATAGCGTCTCCCGGACCGCTCAGTCAACAATGGA 420
421 TACCCGAGCCAAATGCGCGGCGCTCCCGGACCGCTCAGTCAACAATGGA 480
421 TACCCGAGCCAAATGCGCGGCGCTCCCGGACCGCTCAGTCAACAATGGA 480
481 CTTCCTTCAGGAGCATTTCAGTTATTCGGACGGAAGCCTCAGGTTGGACCCCGTTGCT 540
481 CTTCCTTCAGGAGCATTTCAGTTATTCGGACGGAAGCCTCAGGTTGGACCCCGTTGCT 540
541 GCGAATGGTGTGCTGCGCCGATCCCGGGAAGGACTCCCGGACCGGATGGAAGTTGTTCT 600
541 GCGAATGGTGTGCTGCGCCGATCCCGGGAAGGACTCCCGGACCGGATGGAAGTTGTTCT 600
601 CAGGGGACGCGGTATCTTCCGAGTGTTCGGGACGCGCAACTCCGGTCACAAATGGT 660
601 CAGGGGACGCGGTATCTTCCGAGTGTTCGGGACGCGCAACTCCGGTCACAAATGGT 660
661 GTTAACGGCACCGGCAAGAACACTACTATCCCGGCAAGGATGCGGATGGAAGTTCCCT 720
661 GTTAACGGCACCGGCAAGAACACTACTATCCCGGCAAGGATGCGGATGGAAGTTCCCT 720
721 TGCCCGAAGTGTAAAGACTTATCTTCATGCAAGCATCTCAGCGCCATCTGCTACGC 780
721 TGCCCGAAGTGTAAAGACTTATCTTCATGCAAGCATCTCAGCGCCATCTGCTACGC 780
781 CACACTGTGTACCGCCCGTACATGTGTGTTCTTTGCAAGACACCTTCTCTCGCAGTGAT 840
781 CACACTGTGTACCGCCCGTACATGTGTGTTCTTTGCAAGACACCTTCTCTCGCAGTGAT 840
841 ATCCTGAACGTCATTTCCAAAAATGCTCAATCAGGCGTGTGTAACCCACCGGAGCAACG 900
841 ATCCTGAACGTCATTTCCAAAAATGCTCAATCAGGCGTGTGTAACCCACCGGAGCAACG 900

QY 901 CACTTGTGCGACCCCAATCGCATGTGAAGAGGTCCCAACAGCAGGCTCGCGGAATCCT 960
Db 901 CACTTGTGCGACCCCAATCGCATGTGAAGAGGTCCCAACAGCAGGCTCGCGGAATCCT 960
QY 961 GTAAACCTGTCCAGGATGAAGTCACTAGTACCGTCCCGCTCCCAATGGCATCCCGGC 1020
Db 961 GTAAACCTGTCCAGGATGAAGTCACTAGTACCGTCCCGCTCCCAATGGCATCCCGGC 1020
QY 1021 ACGACTTACGGGAGGAGCGGTCAACGGCAATGGACTAGCTCCGGCCCGCCAGGTAC 1080
Db 1021 ACGACTTACGGGAGGAGCGGTCAACGGCAATGGACTAGCTCCGGCCCGCCAGGTAC 1080
QY 1081 GCGGATCACAGACTATGGGTTTCCCAATGTCCGTCAACGGGATGGCGGTGTCAG 1140
Db 1081 GCGGATCACAGACTATGGGTTTCCCAATGTCCGTCAACGGGATGGCGGTGTCAG 1140
QY 1141 CCTGAAGACGGTTTCCCGGCGCGCGCATCAAGGAGCCCTTTGGCCACAAGCTCCC 1200
Db 1141 CCTGAAGACGGTTTCCCGGCGCGCGCATCAAGGAGCCCTTTGGCCACAAGCTCCC 1200
QY 1201 AAGCAGAGCCCGTATCTCGTGAGCCGGTGTGACCCCTTCTGGCCACAGTTGAATATT 1260
Db 1201 AAGCAGAGCCCGTATCTCGTGAGCCGGTGTGACCCCTTCTGGCCACAGTTGAATATT 1260
QY 1261 GACCGAAACATCGAGCAGGTAAACAAACCGGTTGTTCAAGACCCCAAGCGCTGTGATG 1320
Db 1261 GACCGAAACATCGAGCAGGTAAACAAACCGGTTGTTCAAGACCCCAAGCGCTGTGATG 1320
QY 1321 CCAGGACATCCCGGCCACCCCGGTGAGTTGACTGGAGCTCTATGTTCCAACTCAAGCT 1380
Db 1321 CCAGGACATCCCGGCCACCCCGGTGAGTTGACTGGAGCTCTATGTTCCAACTCAAGCT 1380
QY 1381 CCCGAGGCTACATGTTTCTCCAGTCTATGCTGGTGGTCAAGAGCCCAATCCAGCTCAT 1440
Db 1381 CCCGAGGCTACATGTTTCTCCAGTCTATGCTGGTGGTCAAGAGCCCAATCCAGCTCAT 1440
QY 1441 GTCGAGACCGGAGGAAAGTATTACCCCACTACCGTGGTCAAGAGAGTGAATGAAC 1500
Db 1441 GTCGAGACCGGAGGAAAGTATTACCCCACTACCGTGGTCAAGAGAGTGAATGAAC 1500
QY 1501 GGTCTCTATCTGCTTCCAGTATGAGTGGGCGGACCGCTTCCAGCCCGCCAGACAA 1557
Db 1501 GGTCTCTATCTGCTTCCAGTATGAGTGGGCGGACCGCTTCCAGCCCGCCAGACAA 1557

RESULT 2
AF467711
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
gene
AF467711
Emmericella nidulans RfC (rfec) mRNA, complete cds.
1560 bp mRNA linear PLN 01-JAN-2003
Emmericella nidulans
Emmericella nidulans
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; Emmericella.
1 (bases 1 to 1560)
Holtzman, D.A.
Enhanced Production of Secondary Metabolites in Filamentous Fungi
by Ectopic Expression of Regulatory Proteins
Unpublished
2 (bases 1 to 1560)
Holtzman, D.A.
Direct Submission
Submitted (14-JAN-2002) Microbia, Inc., One Kendall Sq. Building
1400W, Cambridge, MA 02139, USA
Location/Qualifiers
1. 1560
/organism="Emmericella nidulans"
/mol_type="mRNA"
/db_xref="taxon:162425"
1. 1560

CDS		/gene="rfec"	
		1..1560	
		/gene="rfec"	
		/note="putative transcription factor; contains zinc finger and C2H2 domain; regulator of F101"	
		/codon_start=1	
		/product="Rf1c"	
		/protein_id="AA014631.1"	
		/db_xref="GI:27450539"	
		/translation="MDPRNHPSPSTSLPQSSAPLPSAPISMPMPQVYTMQPPVPS QPHTLPPLQPHSOPAPHSYMGPPYRPLNRYPASSHDVYASSAPIMPTTVGSL PPTSLSHPNPQAQAQSQSPHYPPHSLVLPASSAQSPQPIAPAPRRDRADFNG LPSGAFSYSDGKPDQWDVPAANGAAPYPGKDSPTQVWVSQRRGILPSVPGRATPVT NGVNGTGKNTIIPAKDADGKFPNPNCKTYLHAKHLKRHLRHTGDRPYMVLCKDTF SRSDILKRHFQKSIIRGNPTGATHLSHPNNAHVRSQQAAANPVKPVQDEVSTVP PNGIPGTTYGEGAVNGNLAPAPRGYADHQTMGFPMSVNGMRGQPEDAFPPGGRPHQ GAPWPAQKQSPYLQPGADPSGHQLNIDRNIHQVQPVQDPKRPVMPGHPGHPGEL DWTSMFQPAPEGYNFSQSMPPGQEPHIAHVETERYKYPYPTTAGQESGMNGLYLASTM SGDGTVPQARQ"	
BASE COUNT		330 a	533 c 386 g 311 t
ORIGIN			
Query Match		100.0%; Score 1557; DB 8; Length 1560;	
Best Local Similarity		100.0%; Pred. No. 0;	
Matches 1557; Conservative		0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	ATGGATCCTAGAAACCATCCTCTCGGCCTCGCTACACAGTCTGCCTCAAGATCGGCG	60
Db	1	ATGGATCCTAGAAACCATCCTCTCGGCCTCGCTACACAGTCTGCCTCAAGATCGGCG	60
QY	61	CCTCTTCTCTGCTCCCATCTCGAGCATGCCAATGCCTCAGTACACGATGACGCTCAG	120
Db	61	CCTCTTCTCTGCTCCCATCTCGAGCATGCCAATGCCTCAGTACACGATGACGCTCAG	120
QY	121	TACCCAGTCTCTCAGCGGCACACCTCGCTCCTCTGCAACCCCATCATAGCCAGTGC	180
Db	121	TACCCAGTCTCTCAGCGGCACACCTCGCTCCTCTGCAACCCCATCATAGCCAGTGC	180
QY	181	GCTCCTCACTCGTACATGGGCGAGCGCGGTACCGGCTGATCTGAACAGGTACCCGCA	240
Db	181	GCTCCTCACTCGTACATGGGCGAGCGCGGTACCGGCTGATCTGAACAGGTACCCGCA	240
QY	241	TCAAGTCACGATGTTTACCGGTCTTCTGCTGGCGCGGTAATGCCCCACACTACCGTGGC	300
Db	241	TCAAGTCACGATGTTTACCGGTCTTCTGCTGGCGCGGTAATGCCCCACACTACCGTGGC	300
QY	301	AGCTTGCCTCCGACATCTTCTCTCATCCCAATCCGAGCGCGAGGCGAGGCGGCGAG	360
Db	301	AGCTTGCCTCCGACATCTTCTCTCATCCCAATCCGAGCGCGAGGCGAGGCGGCGAG	360
QY	361	CAATCGCGCACTATCCTCTCTCATAGCGTGTCTCCCGCGGCTTCCAGCGCTCAGTCG	420
Db	361	CAATCGCGCACTATCCTCTCTCATAGCGTGTCTCCCGCGGCTTCCAGCGCTCAGTCG	420
QY	421	TACCCGCGCACTATGCGCGCGGCGCTCCCGGAGCGGCTCGTGTGACTTCAACAATGGA	480
Db	421	TACCCGCGCACTATGCGCGCGGCGCTCCCGGAGCGGCTCGTGTGACTTCAACAATGGA	480
QY	481	CTTCTCTCAGGAGCATTCAGTTATTTCGGACGGAAGCCTCAAGTTGGACCCCGTTGCT	540
Db	481	CTTCTCTCAGGAGCATTCAGTTATTTCGGACGGAAGCCTCAAGTTGGACCCCGTTGCT	540
QY	541	GCGAATGGTGTGCGCGGTATCCCGGGAAGGACTCCCGCGGAAAGGACTCCCGGAAAGG	600
Db	541	GCGAATGGTGTGCGCGGTATCCCGGGAAGGACTCCCGGAAAGGACTCCCGGAAAGG	600
QY	601	CAGGGCGACGCGGTATCCTTCCGAGTGTTCGGGACGCGCACTCCGCTCACAATGGT	660
Db	601	CAGGGCGACGCGGTATCCTTCCGAGTGTTCGGGACGCGCACTCCGCTCACAATGGT	660
QY	661	GTTAACGGCACCGGCAAGAACACTACTATCCCGGCCAAGGATGCCGATGGAAGTTCCT	720
Db	661	GTTAACGGCACCGGCAAGAACACTACTATCCCGGCCAAGGATGCCGATGGAAGTTCCT	720

QY	721	TGCCCCGAACCTGTAAAGACTTATCTTTCATGCCAAGCATCTCAAGCGCCCATCTGCTACGC	780
Db	721	TGCCCCGAACCTGTAAAGACTTATCTTTCATGCCAAGCATCTCAAGCGCCCATCTGCTACGC	780
QY	781	CACACTGGTGACCGCCCGTACATGTGTGTTCTTTGCAAAAGACACCTTCTCTCGCAGTGAT	840
Db	781	CACACTGGTGACCGCCCGTACATGTGTGTTCTTTGCAAAAGACACCTTCTCTCGCAGTGAT	840
QY	841	ATCCTGAAACCGTCAATTTCCAAAAATGCTCAATCAGGCGTGGTAACCCCAACCGGAGCAACG	900
Db	841	ATCCTGAAACCGTCAATTTCCAAAAATGCTCAATCAGGCGTGGTAACCCCAACCGGAGCAACG	900
QY	901	CACATTGTCGACCCCAATGCGCATGTGAAGAGTCCCAACAGCAGGCTGCGGCAATCCT	960
Db	901	CACATTGTCGACCCCAATGCGCATGTGAAGAGTCCCAACAGCAGGCTGCGGCAATCCT	960
QY	961	GTAACACCTGTCCAGGATGAAGTCAAGTACGTACCGTCCCGCCTCCCAATGCGCATCCCGGC	1020
Db	961	GTAACACCTGTCCAGGATGAAGTCAAGTACCGTCCCGCCTCCCAATGCGCATCCCGGC	1020
QY	1021	ACGACTTACGCGAGGAGCGCGTCAACGSCAATGGACTAGTCTCCGGCCCGCCAGGCTAC	1080
Db	1021	ACGACTTACGCGAGGAGCGCGTCAACGSCAATGGACTAGTCTCCGGCCCGCCAGGCTAC	1080
QY	1081	GCGGATCACCAACTATGGGCTTCCCAATGTTCATCCGTCAACGGGATGGGCGTGGTCAAG	1140
Db	1081	GCGGATCACCAACTATGGGCTTCCCAATGTTCATCCGTCAACGGGATGGGCGTGGTCAAG	1140
QY	1141	CCTGAAGACGCGTTTCCCGGCGCGCGCATCAAGGAGCCCTTTGGCCACCAAGCTCCC	1200
Db	1141	CCTGAAGACGCGTTTCCCGGCGCGCGCATCAAGGAGCCCTTTGGCCACCAAGCTCCC	1200
QY	1201	AAGCAGAGCCCGTATCTCGTGAGCCGGGTGCTGACCCCTTCTGGCCACCAAGTGAATATT	1260
Db	1201	AAGCAGAGCCCGTATCTCGTGAGCCGGGTGCTGACCCCTTCTGGCCACCAAGTGAATATT	1260
QY	1261	GACCGAAACATCGAGCAGGTAAACAAACCGGTTGTTCAAGACCCCAAGCGCCCTGTGATG	1320
Db	1261	GACCGAAACATCGAGCAGGTAAACAAACCGGTTGTTCAAGACCCCAAGCGCCCTGTGATG	1320
QY	1321	CCAGGACATCCCGGCCACCCCGGTGAGTTGACTGGACGCTATGTTTCAACCTCAAGCT	1380
Db	1321	CCAGGACATCCCGGCCACCCCGGTGAGTTGACTGGACGCTATGTTTCAACCTCAAGCT	1380
QY	1381	CCCAGGSGCTACATGTTCTCCAGTCTATGCTGGTGGTCAAGAGCCCATCCACGCTCAT	1440
Db	1381	CCCAGGSGCTACATGTTCTCCAGTCTATGCTGGTGGTCAAGAGCCCATCCACGCTCAT	1440
QY	1441	GTCGAGACCGAGCGAAAGTATTACCCACCACTACCGTGGTCAAGAGAGTGAATGAAC	1500
Db	1441	GTCGAGACCGAGCGAAAGTATTACCCACCACTACCGTGGTCAAGAGAGTGAATGAAC	1500
QY	1501	GGTCTCTATCTGGCTTCGACTATGAGTGGGCGACCGGACCCGTTTCCAGCCCGCAGACAA	1557
Db	1501	GGTCTCTATCTGGCTTCGACTATGAGTGGGCGACCGGACCCGTTTCCAGCCCGCAGACAA	1557

RESULT 3			
AX505252		1800 bp	DNA
LOCUS			
DEFINITION	Sequence 118 from Patent WO02057456.		
ACCESSION	AX505252		
VERSION	AX505252.1	GI:23386501	
KEYWORDS			
SOURCE	synthetic construct		
ORGANISM	synthetic construct		
REFERENCE	artificial sequences.		
AUTHORS	1		
TITLE	Calu, B.M., Madden, K.T., Milne, T.G., Zhang, L., Silva, J.C.,		
JOURNAL	Trueheart, J., Holtzman, D. and Sherman, A.		
	Regulators of fungal gene expression		
	Patent: WO 02057456-A 118 25-JUL-2002;		


```

Microbia (US)
FEATURES             Location/Qualifiers
     source            1. .1800
                        /organism="synthetic construct"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:32630"
                        /note="fungal gene"
BASE COUNT           372 a   606 c   463 g   359 t
ORIGIN
Query Match          100.0%; Score 1557; DB 6; Length 1800;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGATCCTAGAAACCATCCCTCTCGGCCCTCCGTCTACCACTCTGCCTCAAGGATCGGCG 60
    |||
Db 244 ATGGATCCTAGAAACCATCCCTCTCGGCCCTCCGTCTACCACTCTGCCTCAAGGATCGGCG 303

QY 61 CCTCTTCCTTCTGTCTCCCATCTCGAGCATGCCAATGCCTCAGTACACGATGACGCTCAG 120
    |||
Db 304 CCTCTTCCTTCTGTCTCCCATCTCGAGCATGCCAATGCCTCAGTACACGATGACGCTCAG 363

QY 121 TACCCAGTCTCTCAGCGGCACACCCCTGCTCTCTCTGCAACCCCATCATAGCAGTCGCCC 180
    |||
Db 364 TACCCAGTCTCTCAGCGGCACACCCCTGCTCTCTCTGCAACCCCATCATAGCAGTCGCCC 423

QY 181 GCTCCTCACTCGTACATGGGCGAGCCGCCGTACCGGCCTGATCTGAACAGGTACCCCGCA 240
    |||
Db 424 GCTCCTCACTCGTACATGGGCGAGCCGCCGTACCGGCCTGATCTGAACAGGTACCCCGCA 483

QY 241 TCAAGTCACGATGTTTACCGGTCTTCTGTGCGGCCGATAATGCCCCACACTACCGTGGGC 300
    |||
Db 484 TCAAGTCACGATGTTTACCGGTCTTCTGTGCGGCCGATAATGCCCCACACTACCGTGGGC 543

QY 301 AGCTTGCTCCGACATCTTTCCTTTCTCATCCCAATCCGCGAGCGCAGGCAAGGCGCAG 360
    |||
Db 544 AGCTTGCTCCGACATCTTTCCTTTCTCATCCCAATCCGCGAGCGCAGGCAAGGCGCAG 603

QY 361 CAATGCGCGCACTATCCTCCTCTCATAGCGTGTCTCCGCCCCGCTTCCAGCGTCACTCG 420
    |||
Db 604 CAATGCGCGCACTATCCTCCTCTCATAGCGTGTCTCCGCCCCGCTTCCAGCGTCACTCG 663

QY 421 TACCGCGAGCCAAATTGCGCGCGCGCTCCCGGGACCGTCTGCTGACTTCAACAATGGA 480
    |||
Db 664 TACCGCGAGCCAAATTGCGCGCGCGCTCCCGGGACCGTCTGCTGACTTCAACAATGGA 723

QY 481 CTTTCCTTCAGGAGCATTCAGTTATTTCGGACGGAAAGCCTCAAGTTGGGACCCCGTTGCT 540
    |||
Db 724 CTTTCCTTCAGGAGCATTCAGTTATTTCGGACGGAAAGCCTCAAGTTGGGACCCCGTTGCT 783

QY 541 GCGAATGGTGTGTGCGCCGATATCCCGGGAAGGACTCCCGCCGAACCCAGTTGTGTTCT 600
    |||
Db 784 GCGAATGGTGTGTGCGCCGATATCCCGGGAAGGACTCCCGCCGAACCCAGTTGTGTTCT 843

QY 601 CAGGGGCGACGCGGTATCCTTCGAGTGTTCGGGACGCGCAACTCCGGTCAACAATGGT 660
    |||
Db 844 CAGGGGCGACGCGGTATCCTTCGAGTGTTCGGGACGCGCAACTCCGGTCAACAATGGT 903

QY 661 GTTAACGGCACCGGCAAGAACACTACTATCCCGGCCAAGGATGCCGATGGAAGTTCCCT 720
    |||
Db 904 GTTAACGGCACCGGCAAGAACACTACTATCCCGGCCAAGGATGCCGATGGAAGTTCCCT 963

QY 721 TGCCCGAACTGTAAACAGACTTATCTTCATGCCAAGCATCTCAAGCGCCATCTGCTACGC 780
    |||
Db 964 TGCCCGAACTGTAAACAGACTTATCTTCATGCCAAGCATCTCAAGCGCCATCTGCTACGC 1023

QY 781 CACACTGGTGACCGCCCGTACATGTGTGTTCTTTTGCAAGACACCTTCTCTCGCAGTGAT 840
    |||
Db 1024 CACACTGGTGACCGCCCGTACATGTGTGTTCTTTTGCAAGACACCTTCTCTCGCAGTGAT 1083

QY 841 ATCCTGAAACGTCAATTCGAAATGCTCAATCAGGCGTGGTAAACCCACCCGGAGCAACG 900
    |||
Db 1084 ATCCTGAAACGTCAATTCGAAATGCTCAATCAGGCGTGGTAAACCCACCCGGAGCAACG 1143

```

Matches 1557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	ATGGATCCTAGAAACCATCCCTCTCGGCCTCCGTCTACCACTCTGCCTCAAGGATCGCGG	60
Db	277	ATGGATCCTAGAAACCATCCCTCTCGGCCTCCGTCTACCACTCTGCCTCAAGGATCGCGG	336
QY	61	CCTCTCTCTTCTGCTCCCATCTCGAGCATGCCAATGCCCTCAGTACAGATGCAGCTCAG	120
Db	337	CCTCTCTCTTCTGCTCCCATCTCGAGCATGCCAATGCCCTCAGTACAGATGCAGCTCAG	396
QY	121	TACCCAGTCTCTCAGCCGCGACACCCCTGCCCTCTCTGAAACCCCATCATAGCCAGTCGCC	180
Db	397	TACCCAGTCTCTCAGCCGCGACACCCCTGCCCTCTCTGAAACCCCATCATAGCCAGTCGCC	456
QY	181	GCTCCTCACTCGTACATGGGCGAGCCGCCGTACCCGGCTGATCTGAACAGGTACCCCGCA	240
Db	457	GCTCCTCACTCGTACATGGGCGAGCCGCCGTACCCGGCTGATCTGAACAGGTACCCCGCA	516
QY	241	TCAAGTCACGATGTTTACCGCTCTTCTGTGCGCCGATAATGCCACACTACCGTGGC	300
Db	517	TCAAGTCACGATGTTTACCGCTCTTCTGTGCGCCGATAATGCCACACTACCGTGGC	576
QY	301	AGCTTGCTCCGACATCTTCTCTCTCATCCCAATCCGAGCGGCGAGGACAGGCGCAG	360
Db	577	AGCTTGCTCCGACATCTTCTCTCTCATCCCAATCCGAGCGGCGAGGACAGGCGCAG	636
QY	361	CAATCGCGCACTATCTCTCTCTCATAGCGTGTCTCCGCGCCGCTTCCAGGCTCAGTCG	420
Db	637	CAATCGCGCACTATCTCTCTCTCATAGCGTGTCTCCGCGCCGCTTCCAGGCTCAGTCG	696
QY	421	TACCCGCGACCAATTGCGCGCGGCGCTCCCGGAGCCGTCTCGTGTGCTCAACAATGGA	480
Db	697	TACCCGCGACCAATTGCGCGCGGCGCTCCCGGAGCCGTCTCGTGTGCTCAACAATGGA	756
QY	481	CTTCTTTCAGGAGCAATTCAGTTATTTCGGACGGAAGCCTCAAGTTGGGACCCCGTTGCT	540
Db	757	CTTCTTTCAGGAGCAATTCAGTTATTTCGGACGGAAGCCTCAAGTTGGGACCCCGTTGCT	816
QY	541	GCGAATGGTGTGCGCGGTATCCCGGGAAGGACTCCCCCGAACCCAGGTTGTTGTTCT	600
Db	817	GCGAATGGTGTGCGCGGTATCCCGGGAAGGACTCCCCCGAACCCAGGTTGTTGTTCT	876
QY	601	CAGGGCGACCGCGTATCTTCCGAGTGTTCGGGACGCGCAACTCCGTCACAAATGGT	660
Db	877	CAGGGCGACCGCGTATCTTCCGAGTGTTCGGGACGCGCAACTCCGTCACAAATGGT	936
QY	661	GTAAACGGCACCGCAAGAACACTACTATCCCGGCCCAAGGATCCGATGGAAGTTCCCT	720
Db	937	GTAAACGGCACCGCAAGAACACTACTATCCCGGCCCAAGGATCCGATGGAAGTTCCCT	996
QY	721	TGCCCGAAGTGTAAACAGACTTATCTTCATGCGCAAGCATCTCAAGGCGCATCTGTACGC	780
Db	997	TGCCCGAAGTGTAAACAGACTTATCTTCATGCGCAAGCATCTCAAGGCGCATCTGTACGC	1056
QY	781	CACACTGGTGACCGCCCGTACATGTTGTTCTTTGCAAGACACCTTCTCTCGCAGTGAT	840
Db	1057	CACACTGGTGACCGCCCGTACATGTTGTTCTTTGCAAGACACCTTCTCTCGCAGTGAT	1116
QY	841	ATCCTGAAACGTCAATTTCCAAAATGCTCAATCAGGCGTGGTAAACCCACCGGAGCAACG	900
Db	1117	ATCCTGAAACGTCAATTTCCAAAATGCTCAATCAGGCGTGGTAAACCCACCGGAGCAACG	1176
QY	901	CACTTGTGCGACCCCAATGCGCATGTGAAGAGGTCCCAACAGCAGGCTCGCGCAATCCT	960
Db	1177	CACTTGTGCGACCCCAATGCGCATGTGAAGAGGTCCCAACAGCAGGCTCGCGCAATCCT	1236
QY	961	GTAAACCTGTCCAGGATGAAGTCAAGTACCGTCCCGCTCCCAATGGCATCCCGGC	1020
Db	1237	GTAAACCTGTCCAGGATGAAGTCAAGTACCGTCCCGCTCCCAATGGCATCCCGGC	1296
QY	1021	ACGACTACGGCGAGGGAGCCGTCAACGGCAATGGACTAGCTCCGGCCCGCCAGGGTAC	1080
Db	1297	ACGACTACGGCGAGGGAGCCGTCAACGGCAATGGACTAGCTCCGGCCCGCCAGGGTAC	1356
QY	1081	GCGGATCACCAAGACTATGGGCTTCCCAATGTATCCGTCATCCGTCACCGGATGGGCGGTGTCAG	1140
Db	1357	GCGGATCACCAAGACTATGGGCTTCCCAATGTATCCGTCATCCGTCACCGGATGGGCGGTGTCAG	1416
QY	1141	CCTGAAGACGCGTTTCCCGGGCGCGCGCATCAAGGAGCCCTTGGGCCACAAGTCCC	1200
Db	1417	CCTGAAGACGCGTTTCCCGGGCGCGCGCATCAAGGAGCCCTTGGGCCACAAGTCCC	1476
QY	1201	AAGCAGAGCCCGTATCTCGTGACCGCGGTGCTGACCTTCTGGCCACCAAGTGAATATT	1260
Db	1477	AAGCAGAGCCCGTATCTCGTGACCGCGGTGCTGACCTTCTGGCCACCAAGTGAATATT	1536
QY	1261	GACCGAAACATCGAGCAGGTAAACAAACCGGTTGTTCAAGACCCCAAGCGCCTGTGATG	1320
Db	1537	GACCGAAACATCGAGCAGGTAAACAAACCGGTTGTTCAAGACCCCAAGCGCCTGTGATG	1596
QY	1321	CCAGGACATCCCGGCCACCCCGGTGAGTTGACTGGACGTCTATGTTCCCAACCTCAAGCT	1380
Db	1597	CCAGGACATCCCGGCCACCCCGGTGAGTTGACTGGACGTCTATGTTCCCAACCTCAAGCT	1656
QY	1381	CCGAGGCTACATGTTCTCCAGTCTATGCTGCTGCTGCTCAAGAGCCCATCCACGCTCAT	1440
Db	1657	CCGAGGCTACATGTTCTCCAGTCTATGCTGCTGCTGCTCAAGAGCCCATCCACGCTCAT	1716
QY	1441	GTCGAGACCGAGCGAAAGTATTACCCCACTACTCCGCTGGTCAAGAGTGAATGAAC	1500
Db	1717	GTCGAGACCGAGCGAAAGTATTACCCCACTACTCCGCTGGTCAAGAGTGAATGAAC	1776
QY	1501	GGTCTCTATCTGGCTTCGACTATGAGTGGCGACGGCACCGTTCCAGCCCGCCAGACAA	1557
Db	1777	GGTCTCTATCTGGCTTCGACTATGAGTGGCGACGGCACCGTTCCAGCCCGCCAGACAA	1833
RESULT 5			
PM6B12G			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
BASE COUNT			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
QY	601	CAGGGCGGACCGGCTATCCCTTCCGAGTGTTCGGGACGCGCAACTCCGGTCACAAATGGT	660
Db	2	CAGGGACGACGAGGCAATCTTCCAGTGTTCGGGACGCG---CTGCAGTCCCAACGCT	58
QY	661	GTAAACGGCACCGGCAAGAACACTACTATCCCGGCCCAAGGATGCCGATGGAAGTTCCT	720

Db	59	GCCAAATGGAACCGCTAGAACGACCACGATTCGGGCCAAGGATCGCGAGCGCAAGTTCCT	118
QY	721	TGCCCCAACTGTAACAAGACTTATCTTCATGCGCAAGCATCTCAAGCGCATCTGCTACGC	780
Db	119	TGCCCTCATTTGTACCAAGACGTACCTCCACGGAAGCATCTTAAACGCCATCTTTTGAGA	178
QY	781	C-----	ACA 784
Db	179	CGTAAGTTGTTCCACTCAATTCATTACCTTTTCCATCTATTAAACCATCAAAACGCAGACA	238
QY	785	CTGGTGACCGCCGCTACATGTGTGTTCTTTTGCAAAGACACCTTCTCTCGCAGTGATATCC	844
Db	239	CCGGTGATCGCCCTACATGTGTGTTCTCTGCAAAAGATACATTCTCCCGAAGCATATTT	298
QY	845	TGAAACGTCATTTCCAAAAATGCTCAATCAGCGGTGGTAACCCACCGGAGCAACGCACT	904
Db	299	TGAAACGACACTTCCAAAAGTGCTCGTTACGCGCGGTAAACCCACAGGAGCGAGCCATC	358
QY	905	TGTCGCACCCCAATGCGCATGTGAAGAGGTCCCAACAGCAGGCTGGCGGAATCCTGTAA	964
Db	359	TCTCGCATCTCTCAGCGCACCTGAAACGGTC-----TCAGGCCGCGAGCTAATGCTGCAA	412
QY	965	AACCTGTCCAGGATGAAGTCAGTAGTACCGTCCCGCCTCCCAATGGCATCCCGGGCACGA	1024
Db	413	AATCCGTTTCAGGAAGATCAGTAACCTCCATTCCGCTTCCCAATGGTATTGCGGTACGA	472
QY	1025	CTTACGGCGAGGGAGCGCTCAACGGCAATGGACTAGTCTCCGGC-----CCGGCCAG	1075
Db	473	CTTATGGCGACGCGCAGCTGAATGGCCATGGCCACACGCTGGCTCCGGTGGCCGACCTG	532
QY	1076	GGTACGGGATCACCAGACTATGGGCTTCCCAATGTCCATCCGTCAACGGGATGGCCGCTG	1135
Db	533	GTCTCACGGAGCAGCAGCCTCTGGGATATTCAATGTCTCTGCCCCACGGGCTGAACAATG	592
QY	1136	GTCAGCCTGAAGACGGGTTTCCCGGGCGCGCATCAAGGAGCCCTTGGCCACAAG	1195
Db	593	GTCACGTTGGAGACGGCTCAACCCCTGCCAGGAGCATCACAGAGCTTCTGATGGCTG	652
QY	1196	CTCCCAAGCAGAGCCGTATCTCGTGACGCGGGTGCTGACCCCTTCTGGCC	1246
Db	653	CTCAGAAGCAGAGCCCGTACCTTGTNCACTCTTCTGTTCTGCTGAAGGTC	703
RESULT 6			
LOCUS	I66494		
DEFINITION	Sequence 14 from patent US 5670367.	7218 bp	DNA linear PAT 28-DEC-1997
ACCESSION	I66494		
VERSION	I66494.1	GI:2724471	
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unclassified.		
REFERENCE	1 (bases 1 to 7218)		
AUTHORS	Dorner, F., Scheiflinger, F. and Falkner, F. Gunter.		
TITLE	Recombinant fowlpox virus		
JOURNAL	Patent: US 5670367-A 14 23-SEP-1997;		
FEATURES	Location/Qualifiers		
source	1. 7218		
BASE COUNT	1944 a 1491 c 1486 g 1929 t	368	others
ORIGIN			
Query Match	3.9%;	Score 61;	DB 6; Length 7218;
Best Local Similarity	1.9%;	Pred. No. 0.00018;	
Matches	7; Conservative	229; Mismatches	139; Indels 0; Gaps 0;
QY	61	CCTCTTCTCTGCTCCCATCTCGAGCATGCCAATGCTCAGTACACGATGCGCCTCAG	120
Db	1070	YY	1129
QY	121	TACCCAGTCTCTCAGCGGCACACACCTGCTCTCTGCAACCCCATCATAGCCAGTCGCC	180

Db	1130	YY	1189
QY	181	GCTCCTCACTCGTACATGGGCGAGCCCGCTACCGGCTGATCTGAACAGGTACCCCGCA	240
Db	1190	YY	1249
QY	241	TCAAGTCACGATGTTTACGGTCTTCTGCTGGCGCGGATATGCCCCACACTACCGTGGC	300
Db	1250	YY	1309
QY	301	AGCTTGCTCCGACATCTTCTCTTCTCATCCCAATCCGAGGCGCAGGCGGCGAG	360
Db	1310	YY	1369
QY	361	CAATCGCGCACTATCTCTCTCTCATAGCGTGTCTCCGCGCGCTTCCAGCGCTCAGTCG	420
Db	1370	YY	1429
QY	421	TACCCGCGACCAATT	435
Db	1430	YYYYGTACCAATT	1444
RESULT 7			
LOCUS	AC136406/c		
DEFINITION	Rattus norvegicus clone CH230-522A18, *** SEQUENCING IN PROGRESS	241859 bp	DNA linear HTG 23-NOV-2002
ACCESSION	AC136406		
VERSION	AC136406.2	GI:25188350	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.		
SOURCE	Rattus norvegicus (Norway rat)		
ORGANISM	Rattus norvegicus		
REFERENCE	1 (bases 1 to 241859)		
AUTHORS	Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguilano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Chen, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseghe, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L. L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,		

-----Summary Statistics

Center project name: AFC
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 100%
Assembly program: Phrap version 0.990329
Contig length: 219559
Fraction of Phrap value < 40: 0.001207
Error Rate in Consed: 0.01 per 10,000 bases
Number of N's in consensus: 0

----- Distribution of Quality < 40 Bases:

# bases	Phrap Value Range															
	5	10	15	20	25	30	35	40								
1000																
900																
800																
700																
600																
500																
400																
300																
200																
100																
0																

FEATURES

source	Location/Qualifiers
repeat_region	1. 219559
repeat_region	/organism="Mus musculus"
repeat_region	/mol_type="genomic DNA"
repeat_region	/strain="C57BL6/J"
repeat_region	/db_xref="taxon:10090"
repeat_region	/chromosome="5"
repeat_region	/clone="RP23-383N15"
repeat_region	277. 346
repeat_region	/rpt_family="CT-rich"
repeat_region	complement(376. 513)
repeat_region	/rpt_family="ID_B1"
repeat_region	complement(558. 712)
repeat_region	/rpt_family="RSINE1"
repeat_region	complement(1156. 1446)
repeat_region	/rpt_family="ID_B1"
repeat_region	1889. 1981
repeat_region	/rpt_family="B4"
repeat_region	1982. 2201
repeat_region	/rpt_family="B3"
repeat_region	2202. 2384
repeat_region	/rpt_family="B4"
repeat_region	complement(2716. 2856)
repeat_region	/rpt_family="B1F"
repeat_region	complement(2859. 2932)
repeat_region	/rpt_family="ID_B1"
repeat_region	3152. 3288
repeat_region	/rpt_family="WT2B"
repeat_region	3178. 3316
repeat_region	/rpt_family="B1_MM"
repeat_region	3321. 3490
repeat_region	/rpt_family="RSINE1"
repeat_region	3534. 3645
repeat_region	/rpt_family="PB1D10"
repeat_region	3875. 4023
repeat_region	/rpt_family="B1_MM"
repeat_region	4089. 4183
repeat_region	/rpt_family="ZF3AR_MM"
repeat_region	4188. 4270
repeat_region	/rpt_family="ID6"
repeat_region	4795. 4868
repeat_region	/rpt_family="ID_B1"
repeat_region	4809. 4869
repeat_region	/rpt_family="BC1_MM"
repeat_region	4870. 5039
repeat_region	/rpt_family="(GAAA)n"

repeat_region	5040. 5062	/rpt_family="BC1_MM"
repeat_region	complement(6119. 6242)	/rpt_family="PB1D10"
repeat_region	6339. 6482	/rpt_family="RSINE1"
repeat_region	6538. 6593	/rpt_family="(CATATA)n"
repeat_region	complement(6846. 6990)	/rpt_family="ID_B1"
repeat_region	7322. 7385	/rpt_family="A-rich"
repeat_region	complement(7587. 7786)	/rpt_family="MER2"
repeat_region	7870. 8011	/rpt_family="B4A"
repeat_region	9384. 9471	/rpt_family="PB1D10"
repeat_region	9479. 9713	/rpt_family="B4A"
repeat_region	9714. 9816	/rpt_family="PB1D9"
repeat_region	9818. 9840	/rpt_family="(CAAAA)n"
repeat_region	9841. 9865	/rpt_family="B4A"
repeat_region	9866. 9889	/rpt_family="(CAAAA)n"
repeat_region	9890. 9956	/rpt_family="B4A"
repeat_region	10555. 10617	/rpt_family="(CCTG)n"
repeat_region	10731. 11005	/rpt_family="B4A"
repeat_region	11064. 11279	/rpt_family="B4"
repeat_region	11302. 11342	/rpt_family="(CATA)n"
repeat_region	12080. 12111	/rpt_family="AT-rich"
repeat_region	12159. 12228	/rpt_family="GA-rich"
repeat_region	12310. 12499	/rpt_family="B2_Mm1"
repeat_region	12662. 12688	/rpt_family="AT-rich"
repeat_region	12754. 12839	/rpt_family="PB1D7"
repeat_region	13196. 13270	/rpt_family="ID5"

Query Match 3.7%; Score 57.2; DB 10; Length 219559;

Best local Similarity 45.7%; Pred.No. 0.0014;

Matches 200; Conservative 0; Mismatches 238; Indels 0; Gaps 0;

QY	16	CATCCCTCTCGGCTCCGCTACCAAGTCTGCTCAAGGATCGGCGCTCTCTCTCTGCT	75
Db	144415	CCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTG	144474
QY	76	CCCATCTGAGCATGCCAATGCCTCAGTACACGATGAGCCCTCAGTACCCAGTCTCTCAG	135
Db	144475	CCTCCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT	144534
QY	136	CCGCACACCTGCTCTCTGCAACCCCATCATAGCCAGTGGCCGCTCTCTACTCGTAC	195
Db	144535	CTGCTCTCTGCTCTCTGCTCTCTGCTCTCTGCTCTCTGCTCTCTGCTCTCTGCTCT	144594
QY	196	ATGGGGCAGCGCGGTACCGGCTGATCTGAACAGGTACCCCGCATCAAGTACGATGTT	255
Db	144595	CTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT	144654
QY	256	TACGCGTCTTCTGCTGCGCGGATAATGCCACACTACCGTGGGAGCTGCTCCGACA	315

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 2, 2004, 19:09:19 ; Search time 3550 Seconds
(without alignments)
10659.738 Million cell updates/sec

Title: US-10-029-180-7
Perfect score: 1557
Sequence: 1 atggatcctagaaccatcc.....ccgttcagcccgccagacaa 1557

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: em_gss_hum:*
- 18: em_gss_inv:*
- 19: em_gss_pln:*
- 20: em_gss_vrt:*
- 21: em_gss_fun:*
- 22: em_gss_mam:*
- 23: em_gss_mus:*
- 24: em_gss_pro:*
- 25: em_gss_rod:*
- 26: em_gss_phg:*
- 27: em_gss_vrl:*
- 28: gb_gss1:*
- 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	380.4	24.4	442	9 AA785001	AA785001 g4b03a1.r
2	171.6	11.0	768	12 BM077105	BM077105 TrEST-A34
3	108.8	7.0	687	10 BF251517	BF251517 EST418778
4	82.8	5.3	782	28 AQ161501	AQ161501 mgxb00008A

5	82	5.3	585	28	AQ255041	AQ255041 mgxb00008J
6	56.2	3.6	807	13	BX464554	BX464554 BX464554
7	53.8	3.5	1448	29	CC220110	CC220110 CH261-92F
8	52.6	3.4	1201	9	AL515475	AL515475 AL515475
9	52.2	3.4	727	13	BX319543	BX319543 BX319543
10	51.4	3.3	1137	29	CC315110	CC315110 TAM32-30A
11	50.6	3.2	567	28	AZ895651	AZ895651 RPCI-24-2
12	50	3.2	1201	13	BX457568	BX457568 BX457568
13	49.6	3.2	1201	13	BX421672	BX421672 BX421672
14	49.4	3.2	802	29	CNS011T0	AL100686 Drosophil
15	49.2	3.2	786	28	AQ745021	AQ745021 HS 5507 A
16	49.2	3.2	897	29	AG060530	AG060530 Pan trogl
17	49.2	3.2	915	13	BQ952432	BQ952432 AGENCOURT
18	49.2	3.2	1262	12	BM560943	BM560943 AGENCOURT
19	49.2	3.2	1287	10	BF256624	BF256624 HVSMEF001
20	49	3.1	481	9	AA695157	AA695157 GM02329.5
21	49	3.1	759	13	BU314584	BU314584 603543708
22	48.8	3.1	834	28	BZ051898	BZ051898 jnr57e08
23	48.8	3.1	1101	29	CNS000U8	AL051971 Drosophil
24	48.8	3.1	1201	9	AL514195	AL514195 AL514195
25	48.6	3.1	458	10	BF063108	BF063108 7h83e05.x
26	48.6	3.1	648	12	BI227246	BI227246 602949793
27	48.6	3.1	1026	29	AG081192	AG081192 Pan trogl
28	48.4	3.1	732	29	BX149729	BX149729 Danio rer
29	48.4	3.1	815	29	BX239144	BX239144 Danio rer
30	48.4	3.1	987	29	CNS00418	AL066537 Drosophil
31	48.4	3.1	1137	29	AG078502	AG078502 Pan trogl
32	48.2	3.1	786	28	BH018846	BH018846 L2011a.d
33	48.2	3.1	2152	29	CC283954	CC283954 CH261-181
34	48	3.1	785	29	AG081505	AG081505 Pan trogl
35	48	3.1	1130	29	AG060387	AG060387 Pan trogl
36	48	3.1	1201	13	BX356664	BX356664 BX356664
37	47.8	3.1	759	29	AG126157	AG126157 Pan trogl
38	47.8	3.1	774	29	AG128958	AG128958 Pan trogl
39	47.8	3.1	887	29	CNS02SC6	AL211695 Tetraodon
40	47.8	3.1	1082	9	AL574528	AL574528 AL574528
41	47.8	3.1	1104	13	BQ671688	BQ671688 AGENCOURT
42	47.8	3.1	1122	29	CC231036	CC231036 CH261-27F
43	47.6	3.1	731	29	CNS042Q8	AL271817 Tetraodon
44	47.6	3.1	789	14	CA320647	CA320647 UI-M-FW0
45	47.6	3.1	1201	13	BX365323	BX365323 BX365323

ALIGNMENTS

RESULT 1
AA785001
LOCUS

DEFINITION

AA785001 442 bp mRNA linear EST 29-JUL-1998
g4b03a1.r1 Aspergillus nidulans 24hr asexual developmental and
vegetative cDNA lambda zap library Emericella nidulans cDNA clone
g4b03a1 5', mRNA sequence.

ACCESSION

AA785001 GI:2845169

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Emericella nidulans (anamorph: Aspergillus nidulans)
Emericella nidulans
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; Emericella.
1 (bases 1 to 442)
Kupfer, D., Gray, J., Hausner, J., Lai, H., Martin, W., Aramayo, R.,
Prade, R. and Roe, B.
An Aspergillus nidulans EST Database
Unpublished
Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center

Department of Biochemistry
Institute of Chemistry. University of Sao Paulo
Avenida Professor Lineu Prestes, 748, Sao Paulo, SP, 05508-900,
BRASIL
Tel: (55) 11-38183848
Fax: (55) 11-38183848
Email: dorry@iq.usp.br
PCR Primers
FORWARD: Universal M13 forward primer
BACKWARD: Universal M13 reverse primer
Plate: 36 row: F column: 4
Seq primer: M13 reverse primer
High quality sequence stop: 768
POLYA=No.

Location/Qualifiers
1. .768
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM9414 (ATCC26921)"
/db_xref="taxon:51453"
/clone="Tr-A3424"
/sex="Asexual"
/tissue_type="Mycelia"
/dev_stage="18 hr Glycerol-grown culture"
/lab_host="E. coli SOLR cells (kanamycin resistant)"
/clone_lib="TrEST-A"
/note="Vector: pBluescript SK(+); Site 1: EcoRI; Site 2:
XhoI; anamorph=Trichoderma reesei; Cloned unidirectionally
, 5' end of the cDNA cloned into EcoRI site of
pBluescript. Primer: Oligo (dT). Average insert size: 1,2
kb; Uni-ZAP XR Vector system -5' adaptor sequence:
5'GAATTCGGCAGG3' -3' adaptor sequence:
5'CTCGAGTTTCTTTTCTTTTCTTTT3'"

FEATURES
source

E COUNT 92 a 149 c 122 g 79 t
GIN
Query Match 24.4%; Score 380.4; DB 9; Length 442;
est Local Similarity 98.5%; Pred. No. 1.3e-85;
atches 384; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
994 GTCCGGCTCCCAATGGCATCCGGGCACGACTTACGGGAGGGAGCGGTCAACGGCAAT 1053
53 GTCCGGCTCCCAATGGCATCCGGGCACGACTTACGGGAGGGAGCGGTCAACGGCAAT 112
1054 GGACTAGTCCGGCCCGGGCGGATACGGGATCACCGAGACTATGGCTTCCCAATGTCA 1113
113 GGACTAGTCCGGCCCGGGCGGATACGGGATCACCGAGACTATGGCTTCCCAATGTCA 172
1114 TCCGTCAACGGGATGGCGGTGAGCTGAAGACGCGTTTCCCGGCGCGCGCGCAT 1173
173 TCCGTCAACGGGATGGCGGTGAGCTGAAGACGCGTTTCCCGGCGCGCGCGCAT 232
1174 CAAGGAGCCCTTGGCCACAAGCTCCCAAGCAGAGCCCCGTATCTCGTCAGCCGGGTGCT 1233
233 CAAGGAGCCCTTGGCCACAAGCTCCCAAGCAGAGCCCCGTATCTCGTCAGCCGGGTGCT 292
1234 GACCCCTTCTGGCCACCAAGTGAATATTGACCGAATATTCAGGAGGTAACACCGGTT 1293
293 GACCCCTTCTGGCCACCAAGTGAATATTGACCGAATATTCAGGAGGTAACACCGGTT 352
1294 GTTCAAGACCCCAAGCGCCCTGTGATGCCAGGATCCCGGCCACCCCGGTGAGCTTGAC 1353
353 GTTCAAGACCCCAAGCGCCCTGTGATGCCAGGATCCCGGCCACCCCGGTGAGCTTGAC 412
1354 TGGACGTCTATGTTCCAACTCAAGCTCCC 1383
413 TGGACGTCTATGTTCCAACTCAAGCTCCC 442

RESULT 2
4077105
JCUS
EFINITION
BM077105
TrEST-A3424 TrEST-A Hypocrea jecorina cDNA clone Tr-A3424 5', mRNA
sequence.
BM077105
BM077105.1 GI:18498287
JCESSION
ERSON
EYWORDS
SOURCE
ORGANISM
Hypocrea jecorina
Hypocrea jecorina
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreales; Hypocreaceae; Hypocrea.
1 (bases 1 to 768)
Chambergo, F.S., Bonaccorsi, E.D., Ferreira, A.J.S., Ramos, A.S.P.,
Ferreira, J.R.Jr., Abrahao-Neto, J., Farah, J.P.S. and El-Dorry, H.
Elucidation of the metabolic fate of glucose in the filamentous
fungus Trichoderma reesei using expressed sequence tag (EST)
analysis and cDNA microarrays
J. Biol. Chem. 277 (16), 13983-13988 (2002)
21950703
11825887
PUBMED
CONTACT: El-Dorry, Hamza

Seq primer: SK
High quality sequence stop: 429.
Location/Qualifiers
1. .442
/organism="Emericella nidulans"
/mol_type="mRNA"
/strain="FGSC A26"
/db_xref="taxon:162425"
/clone="g4b03a1"
/tissue_type="vegetative mycelia, asexual structures"
/clone_lib="Aspergillus nidulans 24hr asexual
developmental and vegetative cDNA lambda zap library"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
3' end of cDNA cloned into XhoI site of pBluescript"

BASE COUNT 143 a 296 c 191 g 138 t
ORIGIN
Query Match 11.0%; Score 171.6; DB 12; Length 768;
Best Local Similarity 60.7%; Pred. No. 1.2e-32;
Matches 318; Conservative 0; Mismatches 199; Indels 7; Gaps 2;
392 TGCTCCCGCCCGCTTCCAGCGCTCAGTCTGATACCGGAGGCAATGGCGCGCGCTCCCC 451
200 TGATGCCCCAGGCCACCACTGCGGTCTCTCACCCCTCAGCCGATTTGCTCTGCGCGGTTG 259
452 GGGACCGTGTGCTGACTTCAACAATGGACTTCTTTCAGGAGCAATTCAGTTATTCGGACG 511
260 GCGGCGCGGTCTCTCTCTGCTGCTTCCATGCTTCCCGGTGCTCTCATGGCGCAACCCG 319
512 GAAAGCCTCAAGGTTGGGACCCCGTTGCTGCGAATGGTGTGCGCGGTATCCCGGGAAG 571
320 GTGTGGCCTCGCCCTACGGCCCGGGTCTCTGATGCAGCCTAACCCGATCTCCCGGACG 379
572 ACTCCCCCGGAACCCAGGTTGTTGTTCTCAGGGGCGACCGGGTATCTTCCGAGTGTTC 631
380 GCGAGCAGCGCGACCCACGTCGTTGGTTCTCAGGGTCCCGAGGCACTTGGCCAGCGCTC 439
632 CGGACCGCGCAACTCCGCTCACAAATGGTTAAACCGCACCGGCAAGAACACTACTATCC 691
440 CTGGACGGCTACT-----GCCAGGCTGCGGAACTGGCGCAAGAACACCGTCATCC 493
692 CGGCCAAGGATGCCGATGGAAGTTCCCTTTCGCCGAATGTAAACAAGACTTATCTTCATG 751
494 CTGTCAAGGATGCCGACGCGCAATTCCTTTCGCCCTCACTGCACAAAGACGTATCTGCATG 553
752 CCAAGCATCTCAAGCGCCATCTGCTACGCCACACTGGTGACCGCCCGTACATGTGTGTTTC 811
554 CCAAGCATCTCAAGCGCCACCTGCTTTCGCCACACTGGTGATCGCCCTTACATGTGTGTTTC 613
812 TTTGCAAGACACCTTCTCTCGAGTGTATCTTGAACGTCATCTTCCAAAAATGCTCA 870
614 TCTGCCGGGATACCTTCTCCGAGCGACATCTTTCAGGCGGTCTTCCAAAAATGCTTC 673
871 ATCAGGCGGTGTAAACCCACCGGAGCAACGCACTTGTGCGCACCC 914